

University of Massachusetts Amherst

ScholarWorks@UMass Amherst

Masters Theses

Dissertations and Theses

October 2019

Estimating Age-Specific Contraceptive Use for Spacing of Childbirth for All Countries in Sub-Saharan Africa from 1985 to 2030 Using a Bayesian Hierarchical Time Series Model

Gregory Guranich

Follow this and additional works at: https://scholarworks.umass.edu/masters_theses_2



Part of the [Women's Health Commons](#)

Recommended Citation

Guranich, Gregory, "Estimating Age-Specific Contraceptive Use for Spacing of Childbirth for All Countries in Sub-Saharan Africa from 1985 to 2030 Using a Bayesian Hierarchical Time Series Model" (2019).

Masters Theses. 833.

https://scholarworks.umass.edu/masters_theses_2/833

This Open Access Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

**ESTIMATING AGE-SPECIFIC CONTRACEPTIVE USE
FOR SPACING OF CHILDBIRTH FOR ALL COUNTRIES
IN SUB-SAHARAN AFRICA FROM 1985 TO 2030 USING
A BAYESIAN HIERARCHICAL TIME SERIES MODEL**

A Thesis Presented

by

GREGORY B GURANICH

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

September 2019

Department of Biostatistics and Epidemiology

Copyright © 2019 by Gregory B Guranich
All Rights Reserved

**ESTIMATING AGE-SPECIFIC CONTRACEPTIVE USE
FOR SPACING OF CHILDBIRTH FOR ALL COUNTRIES
IN SUB-SAHARAN AFRICA FROM 1985 TO 2030 USING
A BAYESIAN HIERARCHICAL TIME SERIES MODEL**

A Thesis Presented

by

GREGORY B GURANICH

Approved as to style and content by:

Leontine Alkema, Chair

Krista Gile, Member

Laura Balzer, Member

Lisa Chasan-Taber, Department Chair
Department of Biostatistics and Epidemiology

ACKNOWLEDGMENTS

We thank the UN populations division, in particular Philipp Ueffing and Vladimira Kantarova for providing us with compiled data.

This research is funded by the Bill and Melinda Gates Foundation.

ABSTRACT

ESTIMATING AGE-SPECIFIC CONTRACEPTIVE USE FOR SPACING OF CHILDBIRTH FOR ALL COUNTRIES IN SUB-SAHARAN AFRICA FROM 1985 TO 2030 USING A BAYESIAN HIERARCHICAL TIME SERIES MODEL

SEPTEMBER 2019

GREGORY B GURANICH

B.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Leontine Alkema

Contraceptive usage for spacing of childbirth is an important indicator for understanding family planning practices as well as fertility transitions. Fertility transitions are especially important in sub-Saharan Africa where fertility remains high in many countries. However, estimates and short-term projections are generally not available for countries in this region. We developed a Bayesian hierarchical time series model to estimate and project usage of contraceptives for spacing by 5-year age groups for all countries in sub-Saharan Africa for the years 1985-2030. Estimating country-age-

year specific usage is challenging due to limited data availability. We use Bayesian hierarchical models to share information across countries and spline regression to share information across age groups. Temporal changes are captured with logistic growth curves and autocorrelated distortion terms. Models are validated with out of sample exercises which test the model’s ability to project into the future as well as the models ability to estimate historical trends. Validation results show the model is well calibrated. Estimates reveal noteworthy variability across countries and across age groups.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
ABSTRACT	v
LIST OF TABLES	ix
LIST OF FIGURES	x
 CHAPTER	
1. INTRODUCTION	1
2. METHODS	3
2.1 Data	3
2.1.1 Indicator	3
2.1.2 Database	3
2.1.3 Exploratory data analysis	4
2.2 Model overview	4
2.3 Technical specifications of model	9
2.3.1 Data model	9
2.3.2 Process model	10
2.4 Model details	11
2.4.1 Logistic growth	11

2.4.2	B-splines	13
2.4.3	Bayesian hierarchical models	15
2.4.4	Full model	16
2.5	Computation	18
2.6	Validation	19
3.	RESULTS	20
3.1	Country estimates and projections	20
3.1.1	Illustrative country estimates and projections	20
3.1.2	Choropleth maps	20
3.2	Model parameters	22
3.2.1	Asymptotes	22
3.2.2	Pace of contraceptive transition parameters	29
3.3	Validation Results	32
4.	DISCUSSION	33
	APPENDIX: SUPPLEMENTARY INFORMATION	35
	BIBLIOGRAPHY	86

LIST OF TABLES

Table	Page
3.1 Validation results I: measures of absolute error	32
3.2 Validation results II: measures of error	32
3.3 Validation results III: coverage of left-out observations	32

LIST OF FIGURES

Figure		Page
2.1	Exploratory plots: spacing \sim age	5
2.2	Exploratory plots: spacing \sim age (by region)	6
2.3	Exploratory plots: spacing \sim time	7
2.4	Exploratory plots: spacing \sim time (by region and age)	8
2.5	Spline placement	14
2.6	Illustrative spline model	15
3.1	Country estimates and projections 1985-2030 by 5-year age groups	21
3.2	Estimates of spacing for the year 2019, ages 15-34	23
3.3	Estimates of spacing for the year 2019, ages 35-49	24
3.4	Estimate change in spacing (2000-2019), ages 15-34	25
3.5	Estimate change in spacing (2000-2019), ages 35-49	26
3.6	Regional asymptotes	27
3.7	Country asymptotes — Eastern Africa	28
3.8	Regional pace parameters	30

3.9	Country pace parameters	31
A.1	Country asymptotes by region	36
A.2	Country estimates and projections 1985-2030 by 5-year age groups	39

CHAPTER 1

INTRODUCTION

The right to health includes freely deciding whether, when, and how many children to have. In 2012, 20 national governments came together at the London summit to plan a global movement to address this right. The goal of the movement is to address policy, financing, delivery, and socio-cultural barriers to accessing contraceptive information, services, and supplies. The product of this summit is the FP2020 initiative where the goal is to reach 120 million additional users of modern contraceptive methods by 2020 (Brown et al. 2014). These goals are also in alignment with Sustainable Development Goals (SDGs) goals for good health (goal 3) and well-being and gender and equality (goal 5).

The Family Planning Estimation Tool (FPET) (JR New) is used to track countries' progress toward the FP2020 goal. It is a modeling approach that provides estimates and projections for family planning indicators (Alkema et al. 2013; Cahill et al. 2018). The tool is used by Track 20 and the UN population division.

Contraceptive usage for spacing of childbirth is an important indicator for understanding family planning practices as well as fertility transitions. Fertility transitions are especially important in sub-Saharan Africa where fertility remains high in many countries. Estimates and short-term projections are not produced by the current

FPET model.

In this study, we augment current FPET methods, modeling prevalence of spacing in 5-year age groups. The augmented model is presented in chapter 2 followed by results presented in chapter 3. The model gives us short term projections and estimates of historical trends.

CHAPTER 2

METHODS

2.1 Data

2.1.1 Indicator

Contraceptive usage is a measure of the percentage of self-reported use of at least one type of contraceptive (modern or traditional) by women or by their partners for women who are married or in union. Contraceptive usage for spacing is the subset of these families who not wish to have children for 2 or more years (DHS 2006).

2.1.2 Database

The UN population division (UNPD) compiles data on family planning (United Nations and Social Affairs 2017). The database used in this study was compiled by the UNPD and is composed of nationally representative household surveys, the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS). Our aggregate indicator is the prevalence of contraceptive use for spacing by 5-year age groups within a country for a specific year. The age groups are: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49. Aggregates include nationally representative sampling weights which take into account the two-stage sampling design of the surveys. In total, there are 289 survey campaigns from 48 countries in sub-Saharan

Africa, which provide us with 983 observations of country-age-time specific contraceptive prevalence between 1990 and 2016.

2.1.3 Exploratory data analysis

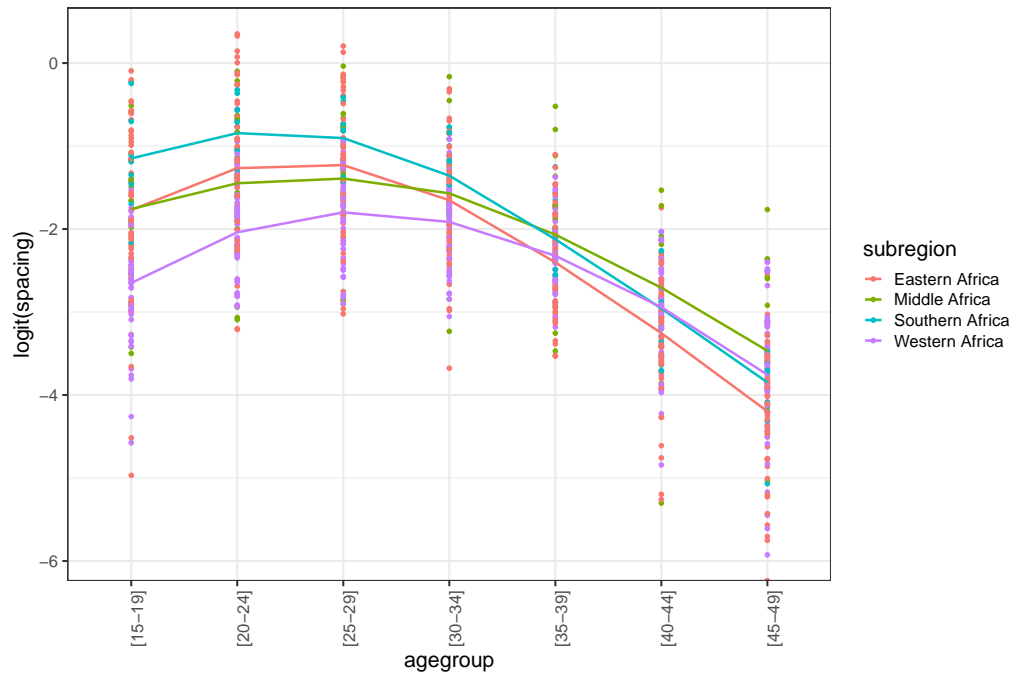
For analysis, we consider age since we expect substantive differences in spacing between age groups. When we plot spacing against age, we see a curve which initially increases and peaks around ages 25-29 followed by a decrease through ages 45-49 (see figure 2.1). There is some variation among regions. For example, Southern Africa peaks early. Western Africa is relatively flat, with a lower starting point, followed by a gradual decrease. Within regions, we see similar trends with some deviations from this regional pattern (see figure 2.2). Spacing over time tends to increase gradually. It is unclear whether the increase depends on the age group (see figure 2.3 and 2.4).

2.2 Model overview

We develop a Bayesian hierarchical time series model to estimate and project usage of contraceptives for spacing by 5-year age groups for all countries in sub-Saharan Africa. The model consists of two major components: the level in the reference year, and a temporal trend. Getting estimates for age groups within countries presents a challenge given limited data. We discuss our approach to this limitation below.

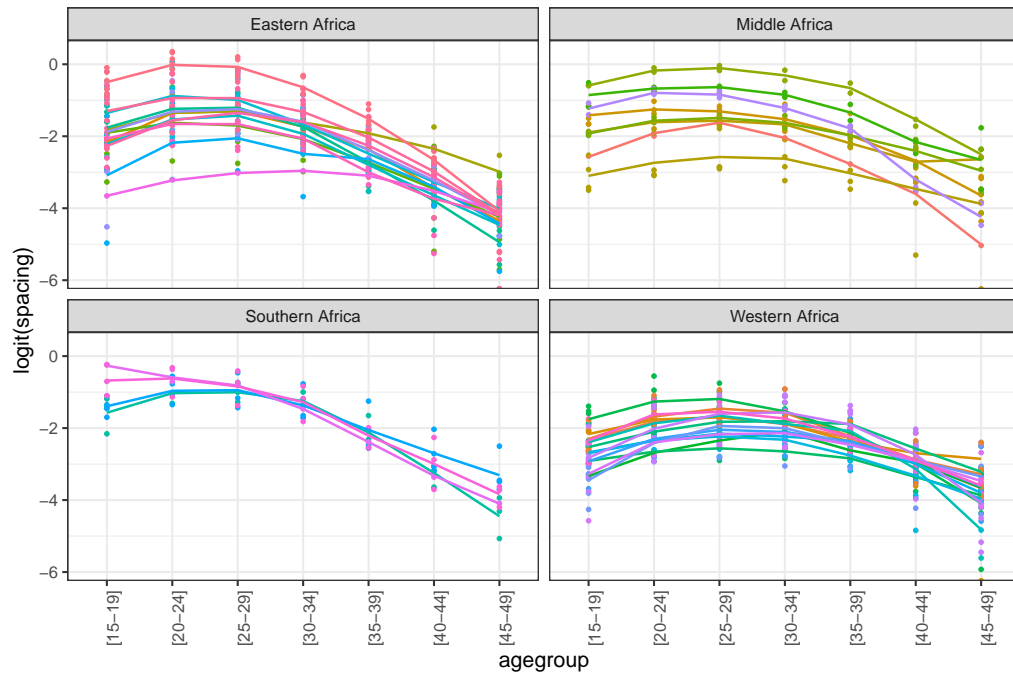
The level in the reference year is modeled with Bayesian hierarchical B-splines, which captures the relationship between age groups and spacing with a smooth, semi-parametric curve. We refer to this relationship between the age group and spacing as the age schedule. The B-spline allows us to share information across age groups.

Figure 2.1. Exploratory plots: spacing \sim age



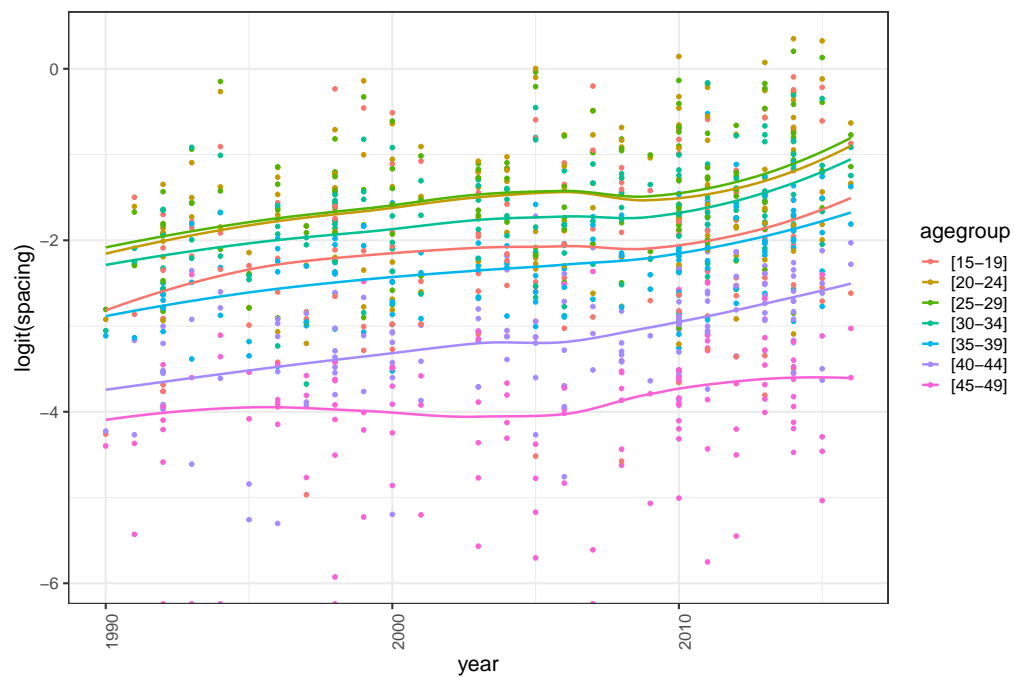
Dots represent observed data on the logit scale from sub-Saharan Africa. Lines fit with LOWESS (locally weighted scatterplot smoothing) curves.

Figure 2.2. Exploratory plots: spacing \sim age (by region)



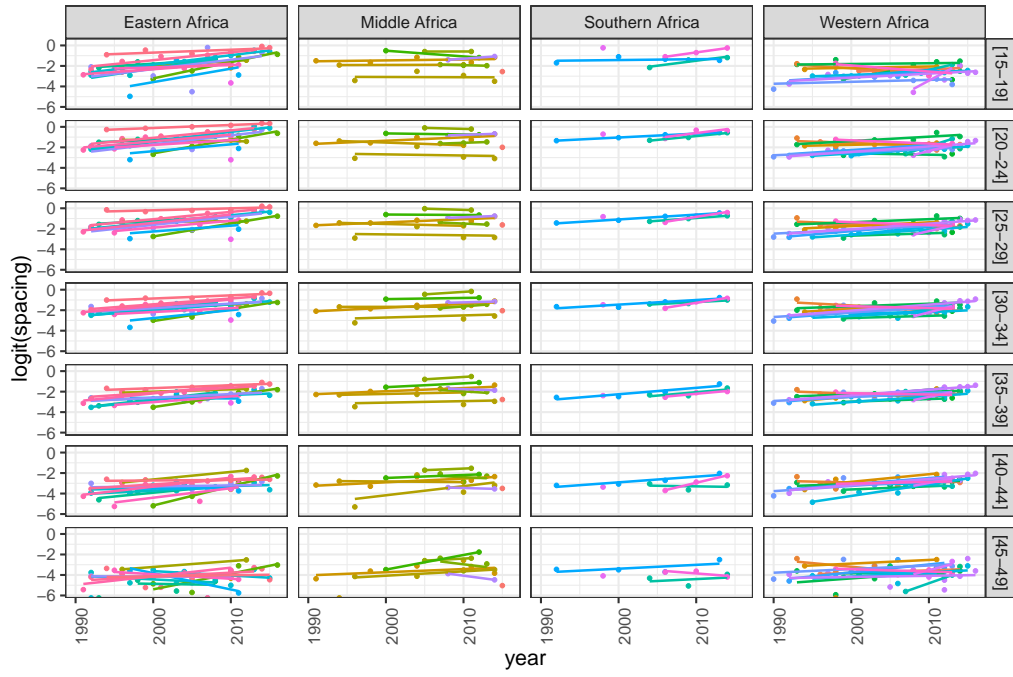
Dots represent observed data on the logit scale from sub-Saharan Africa. Lines fit with LOWESS (locally weighted scatterplot smoothing) curves.

Figure 2.3. Exploratory plots: spacing \sim time



Dots represent observed data on the logit scale from sub-Saharan Africa. Lines were obtained from LOWESS (locally weighted scatterplot smoothing) curves.

Figure 2.4. Exploratory plots: spacing \sim time (by region and age)



Dots represent observed data on the logit scale from sub-Saharan Africa. Lines were obtained from linear models.

Country-specific parameters for the splines are modeled hierarchically, using country information as well as information from the respective sub-region and global region.

Country-age-specific time trends are expected to be increasing and modeled with a logistic growth curve (Alkema et al. 2013; Cahill et al. 2018). Logistic curves start with a gradual increase, followed by a transition in which the increase is rapid, followed by a slow down as high levels of spacing are reached. Estimates are allowed to deviate from this growth curve if data suggest a faster or slower increase than indicated by the logistic growth curve. We model parameters for this expected increase hierarchically. The structure incorporates information from specific country-age groups as well as information from the respective sub-region-age group.

2.3 Technical specifications of model

The model consist of a data model and a process model. The process model is our description of the truth. The data model links our description of the truth to the real world data.

2.3.1 Data model

Observed country-age-year-specific prevalence of contraceptive spacing y_i , is linked to our model of the truth as follows:

$$y_i | \eta_{a[i],c[i],t[i]}, \tau_i^2 \sim N(\eta_{a[i],c[i],t[i]}, \tau_i^2) T(0, 1),$$

where y_i are independent given the estimated age-country-year specific prevalence and variance τ_i^2 . Observed prevalence will be between 0 and 1. Thus the distribution of observed prevalence is truncated. Total error variance τ_y^2 is the sum of the sampling variance s_i^2 and the model estimated non-sampling variance σ_y^2 .

$$\tau_i^2 = s_i^2 + \sigma_y^2.$$

In the current model the sampling error $s_i = \sqrt{\frac{y_i(1-y_i)}{n_i}}$ where n_i refers to the number of women in the sample. This expression is based on the incorrect assumption of having independent samples (see chapter 4, discussion).

2.3.2 Process model

The goal is to model $\eta_{a,c,t}$, age-country-year specific prevalence of contraceptive spacing. The foundation for capturing systematic trends in contraceptive prevalence is a logistic curve scaled with an asymptote (Alkema et al. 2013; Cahill et al. 2018). The following is an example of a logistic growth curve $\eta^*_{a,c,t}$ where

$$\eta^*_{a,c,t} = \frac{P}{1 + e^{-\omega(t-\Omega)}},$$

with asymptote P , pace parameter ω , year t , and t-value at the sigmoid's midpoint Ω . The realized curve, $\eta_{a,c,t}$, may deviate away from this logistic growth curve. We allow for that by setting,

$$\text{logit}(\eta_{a,c,t}) = \text{logit}(\eta_{a,c,t-1}) + \delta_{a,c,t} + \epsilon_{a,c,t}, \quad (1)$$

where $\delta_{a,c,t}$ is some expected change and $\epsilon_{a,c,t}$ is the distortion which allows the model to deviate away from its expected growth. This is explained in more in detail in section 2.4.1.

To complete the specifications for logistic growth we need to specify the distribution of $\eta_{a,c,t=t_0}$, the usage of spacing in the reference year t_0 . We set $t_0 = 2003$ which is the midpoint of the observation period in the data base. The reference year model is explained in detail in section 2.4.2.

Given data limitations we exchange information across age groups and across countries. Information exchange across age groups with basis splines (B-splines) is explained in section 2.4.2. Information exchange across countries with Bayesian hierarchical models is explained in section 2.4.3

2.4 Model details

2.4.1 Logistic growth

To capture the increase in usage over time, we use a logistic growth curve where

$$\eta_{a,c,t}^* = \frac{1}{1 + e^{-\omega_{a,c}(t-\Omega)}},$$

with country-age specific pace parameter $\omega_{a,c}$, year t , and t-value at the sigmoid's midpoint Ω . In this simplified setting where the asymptote equals 1, the prevalence can be expressed where prevalence at time t is dependent on $t - 1$ as follows

$$\text{logit}(\eta_{a,c,t}^*) - \text{logit}(\eta_{a,c,t-1}^*) = \omega_{a,c}.$$

However, spacing within a country-age group is expected to stop increasing some time after a contraceptive transition. We model this slow down and eventual halt with an asymptote as follows.

$$\eta^{*}_{a,c,t} = \frac{P_{a,c}}{1 + e^{-\omega_{a,c}(t-\Omega)}}.$$

Country-age specific asymptotes $P_{a,c}$ are modeled with age schedules. The model for asymptotes is explained in more detail in section 2.4.2. We expect asymptotes to be less than 1.

$$\text{logit}\left(\frac{\eta^{*}_{a,c,t}}{P_{a,c}}\right) - \text{logit}\left(\frac{\eta^{*}_{a,c,t-1}}{P_{a,c}}\right) = \omega_{a,c}.$$

Rewriting the expression above for $\eta^{*}_{a,c,t}$

$$\eta^{*}_{a,c,t} = P_{a,c} * \text{logit}^{-1}\left(\text{logit}\left(\frac{\eta^{*}_{a,c,t-1}}{P_{a,c}}\right) + \omega_{a,c}\right) \quad \text{for } \eta^{*}_{a,c,t-1} < P_{a,c}. \quad (2)$$

Eq 1. in the absence of distortion can be written as

$$\text{logit}(\eta^{*}_{a,c,t}) = \text{logit}(\eta^{*}_{a,c,t-1}) + \delta_{a,c,t}.$$

$$\delta_{a,c,t} = \begin{cases} \text{logit}(P_{a,c} \cdot \text{logit}^{-1}(\text{logit}(\frac{\eta^{*}_{a,c,t-1}}{P_{a,c}}) + \omega_{a,c})) - \text{logit}(\eta^{*}_{a,c,t-1}) & \text{for } \eta_{a,c,t-1} < P_{a,c}, \\ 0 & \text{otherwise.} \end{cases}$$

In reality, we expect deviations away from this expected increase as in Eq. 1. Replacing $\eta_{a,c,t}^*$ with $\eta_{a,c,t}$ and adding distortion we get

$$\begin{aligned} \text{logit}(\eta_{a,c,t}) &= \text{logit}(\eta_{a,c,t-1}) + \delta_{a,c,t} + \epsilon_{a,c,t} \\ &= \begin{cases} \text{logit}(P_{a,c} \cdot \text{logit}^{-1}(\text{logit}(\frac{\eta_{a,c,t-1}}{P_{a,c}}) + \omega_{a,c})) + \epsilon_{a,c,t} & \text{for } \eta_{a,c,t-1} < P_{a,c}, \\ \text{logit}(\eta_{a,c,t-1}) + \epsilon_{a,c,t} & \text{otherwise.} \end{cases} \end{aligned}$$

We estimate spacing from 1985-2030 with forward and backward extrapolation from $t_0 = 2003$. Deviations away from the expected increase for a particular year, written as $\epsilon_{a,c,t}$, are captured with an autoregressive time series process of order 1 (AR(1) process):

$$\epsilon_{a,c,t} | \epsilon_{a,c,t-1}, \rho, \sigma^2 \sim N(\rho \cdot \epsilon_{a,c,t-1}, \sigma^2)$$

The expression above is used for forward extrapolation. Similarly, we can write backwards extrapolation as,

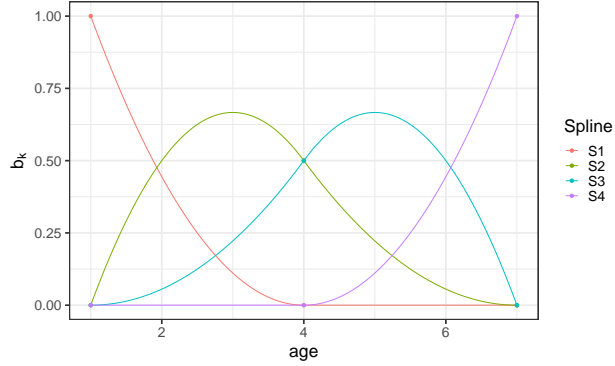
$$\begin{aligned} \text{logit}(\eta_{a,c,t}) &= \text{logit}(\eta_{a,c,t+1}) - \delta_{a,c,t} - \epsilon_{a,c,t}, \\ &= \begin{cases} \text{logit}(P_{a,c} \cdot \text{logit}^{-1}(\text{logit}(\frac{\text{logit}^{-1}(\text{logit}(\eta_{a,c,t+1}) - \epsilon_{a,c,t})}{P_{a,c}}) - \omega_{a,c})) & \text{for } \text{logit}^{-1}(\text{logit}(\eta_{a,c,t+1}) - \epsilon_{a,c,t}) < P_{a,c}, \\ \text{logit}(\eta_{a,c,t+1}) - \epsilon_{a,c,t} & \text{otherwise.} \end{cases} \end{aligned}$$

2.4.2 B-splines

The relationship between spacing and age, which we will refer to as the age-schedule, is modeled with Bayesian B-splines (Eilers and Marx 1996).

We use $K = 4$ splines to smooth over 7 age groups (see figure 2.5). The spline function is a piecewise polynomial function. The places where the pieces meet are

Figure 2.5. Spline placement



known as knots. Spline knots are equidistant with the central knot placed at $a = 4$, corresponding to the fourth age group (ages "30-34"). The placement of the remaining two knots follows. In this set up, we are able to share information across age groups.

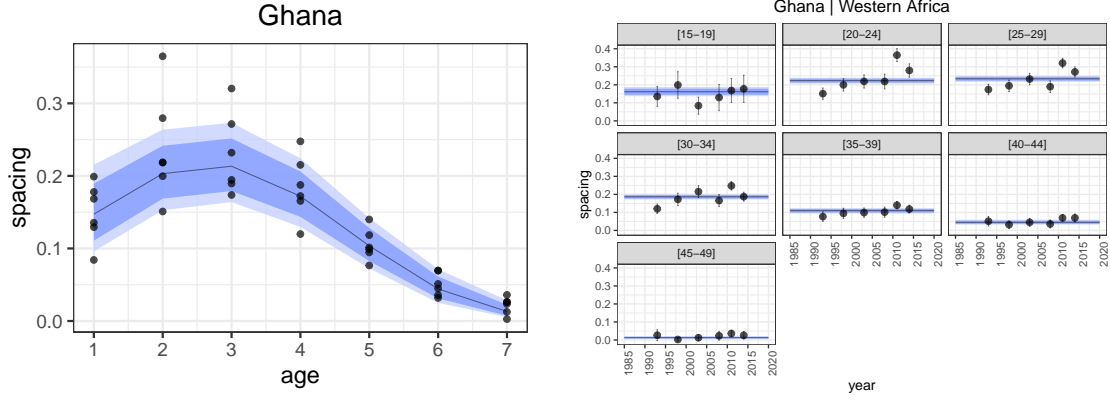
These splines provide us with country-age-specific estimates of the level of spacing. One use for this method is to obtain the level in the reference year (see figure 2.6). The age-schedule for the reference year is written as

$$\eta_{a,c,t_0} = \frac{P_{a,c}}{1 + \exp\left(\underbrace{\sum_{k=1}^K b_k(a)\alpha_{k,c}}_{\text{country-age schedule}}\right)},$$

where b_k is the k -th B-spline evaluated at a and $\alpha_{c,k}$ is the spline coefficient for the country-specific k 'th spline.

Another use for this method is to obtain the maximum level of spacing for a country-age group. Similarly, $\text{logit}(P_{a,c}) = \sum_{k=1}^K b_k(a)\alpha'_{k,c}$.

Figure 2.6. Illustrative spline model



Country-age specific levels of spacing modeled with cubic B-splines. The simplified model $\text{logit}(\eta_{a[i],c[i],t[i]}) = \sum_{k=1}^K b_k(a)\alpha_{k,c}$ was fit without a time trend to illustrate the component. The light blue shaded area represents the 95% credible interval and the dark blue shaded area represents the 90% credible interval. Black dots are observed data with error bars representing sampling variance.

2.4.3 Bayesian hierarchical models

Estimating country-age-specific parameters presents a challenge due to the small number of annual aggregate observations per country age group. We use Bayesian hierarchical models such that estimates come from information within a country as well as information from the respective sub-region (Lindley and Smith 1972; Gelman and Hill 2006).

We use different levels of hierarchy for different portions of the model to accommodate differences in estimation. Both the asymptotes and levels, modeled by B-splines, vary across regions and across countries within regions. The distributions for spline coefficients are as follows:

$$\alpha_{k,c} | \alpha_{k,r[c]}, \sigma_{\alpha_c}^2 \sim N(\alpha_{k,r[c]}, \sigma_{\alpha_c}^2) \quad k = 1, \dots, K,$$

$$\alpha_{k,r[c]} | \alpha_k, \sigma_{\alpha_r}^2 \sim N(\alpha_k, \sigma_{\alpha_r}^2) \quad k = 1, \dots, K,$$

$$\alpha_k \sim N(0, 100) \quad k = 1, \dots, K,$$

with $\alpha_{c,k}$ centered at the sub-regional mean $\alpha_{r[c],k}$ ($r[c]$ refers to the sub-region r where country c resides) with variance across countries σ_{α} . In the case where some country c has no data, the mean of $\alpha_{c,k}$ is equal to the mean of $\alpha_{r[c],k}$ with additional variance given by $\sigma_{\alpha_c}^2$. For countries with data, the estimate is a weighted average of the regional mean and country-specific information regarding the parameter. A simplified model has been fit for illustration (see figure 2.6).

Pace parameters $\omega_{a,c}$ are also expected to vary across regions and across countries. The distribution is as follows:

$$\omega_{a,c}^* = \log\left(\frac{\omega_{a,c} - 0.01}{0.5 - \omega_{a,c}}\right)$$

$$\omega_{a,c} | \omega_{a,r[c]}, \sigma_{\omega}^{2(c)} \sim N(\omega_{a,r[c]}, \sigma_{\omega}^{2(c)})$$

$$\omega_{a,r[c]} | \omega_{r[c]}, \sigma_{\omega}^{2(a,r)} \sim N(\omega_{r[c]}, \sigma_{\omega}^{2(a,r)})$$

$$\omega_{r[c]} | \omega, \sigma_{\omega}^{2(r)} \sim N(\omega, \sigma_{\omega}^{2(r)})$$

$$\omega \sim N(0, 100)$$

such that logit-transformed $\omega_{a,c}$ is restricted between 0.01 and 0.5. This range corresponds to the transition between 10% and 90% of $P_{a,c}$ taking at least 10 years, and a maximum of 4 centuries.

2.4.4 Full model

Data model

$$Y_i | \eta_{a[i],c[i],t[i]}, \tau_y^2 \sim N(\eta_{a[i],c[i],t[i]}, \tau_y^2) T(0, 1)$$

$$\tau_y^2 = s_i^2 + \sigma_y^2$$

Process model

$$\text{logit}\left(\frac{\overbrace{\eta_{a,c,t=t_0}}^{\text{reference year}}}{P_{a,c}}\right) = \underbrace{\sum_{k=1}^K b_k(a) \alpha_{k,c}}_{\text{age schedule}}$$

$$\underbrace{\text{logit}(\eta_{a,c,t>t_0})}_{\text{forward extrapolation}} = \begin{cases} \text{logit}(P_{a,c} \cdot \text{logit}^{-1}(\text{logit}(\frac{\eta_{a,c,t-1}}{P_{a,c}}) + \omega_{a,c}^*)) + \epsilon_{a,c,t} & \text{for } \eta_{a,c,t-1} < P_{a,c} \\ \text{logit}(\eta_{a,c,t-1}) + \epsilon_{a,c,t} & \text{otherwise} \end{cases}$$

$$\underbrace{\text{logit}(\eta_{a,c,t<t_0})}_{\text{backwards extrapolation}} = \begin{cases} \text{logit}(P_{a,c} \cdot \text{logit}^{-1}(\text{logit}(\frac{\text{logit}^{-1}(\text{logit}(\eta_{a,c,t+1}) - \epsilon_{a,c,t})}{P_{a,c}}) - \omega_{a,c}^*)) & \text{for } \eta_{a,c,t-1} < P_{a,c} \\ \text{logit}(\eta_{a,c,t+1}) - \epsilon_{a,c,t} & \text{otherwise} \end{cases}$$

$$\text{logit}(P_{a,c}) = \sum_{k=1}^K b_k(a) \alpha'_{k,c}$$

$$\epsilon_{a,c,t=t_0} | \sigma_\epsilon^2, \rho^2 \sim N(0, \frac{\sigma_\epsilon^2}{1 - \rho^2})$$

$$\epsilon_{a,c,t>t_0} | \rho, \epsilon_{a,c,t-1}, \sigma_\epsilon^2 \sim N(\rho \cdot \epsilon_{a,c,t-1}, \sigma_\epsilon^2)$$

$$\epsilon_{a,c,t<t_0} | \rho, \epsilon_{a,c,t+1}, \sigma_\epsilon^2 \sim N(\rho \cdot \epsilon_{a,c,t+1}, \sigma_\epsilon^2)$$

Prior distributions

$$\alpha_{k,c} | \alpha_{k,r[c]}, \sigma_{\alpha_c}^2 \sim N(\alpha_{k,r[c]}, \sigma_{\alpha_c}^2) \quad k = 1, \dots, K,$$

$$\alpha_{k,r[c]} | \alpha_k, \sigma_{\alpha_r}^2 \sim N(\alpha_k, \sigma_{\alpha_r}^2) \quad k = 1, \dots, K,$$

$$\alpha_k \sim N(0, 100) \quad k = 1, \dots, K,$$

$$\alpha'_{k,c} | \alpha'_{k,r[c]}, \sigma_{\alpha'_c}^2 \sim N(\alpha'_{k,r[c]}, \sigma_{\alpha'_c}^2) \quad k = 1, \dots, K$$

$$\alpha'_{k,r[c]} | \alpha'_k, \sigma_{\alpha'_r}^2 \sim N(\alpha'_k, \sigma_{\alpha'_r}^2) \quad k = 1, \dots, K$$

$$\alpha'_k \sim N(0, 100)$$

$$\omega_{a,c}^* = \log\left(\frac{\omega_{a,c} - 0.01}{0.5 - \omega_{a,c}}\right)$$

$$\omega_{a,c} | \omega_{a,r[c]}, \sigma_{\omega}^{2(c)} \sim N(\omega_{a,r[c]}, \sigma_{\omega}^{2(c)})$$

$$\omega_{a,r[c]} | \omega_{r[c]}, \sigma_{\omega}^{2(a,r)} \sim N(\omega_{r[c]}, \sigma_{\omega}^{2(a,r)})$$

$$\omega_{r[c]} | \omega, \sigma_{\omega}^{2(r)} \sim N(\omega, \sigma_{\omega}^{2(r)})$$

$$\omega \sim N(0, 100)$$

$$\sigma_y, \sigma_{\alpha_c}, \sigma_{\alpha_r}, \sigma_{\alpha'_c}, \sigma_{\alpha'_r}, \sigma_{\epsilon}, \sigma_{\omega}^{(c)}, \sigma_{\omega}^{(a,r)}, \sigma_{\omega}^{(r)} \sim U(0, 2)$$

$$\rho \sim U(0, 1)$$

2.5 Computation

Samples from posterior distributions of model parameters are obtained using Markov chain Monte Carlo (MCMC) algorithm, implemented in open-source software packages R 3.0 (R Core Team 2018) and JAGS 3.2.0 (Plummer et al. 2003), using R-packages R2jags (Su and Yajima 2015) and rjags (Plummer 2018). We obtain results from 6 chains: the total number of iterations in each chain is 26,000, the first 6,000 discarded as burn-in, and after thinning we retain 1,000 samples from each chain. We check the convergence of the MCMC algorithm and sufficiency of the number

of samples through visual inspection of trace plots and convergence diagnostics of Gelman and Rubin (Gelman and Rubin 1992), implemented in the coda R-package (Plummer et al. 2006).

2.6 Validation

Two different procedures were used to validate our model. First, the most recent observation was left-out, and the rest were used for training (Out-of-sample 1). The model’s ability to estimate historical trends while more recent observations are available is of importance as well. Thus our second training set was formed by leaving out 20% of observations at random (Out-of-sample 2).

For left out observations, errors are defined as $e_j = y_j - \hat{y}_j$ where j denotes the left-out index, y_j is the observed prevalence and \hat{y}_j is the posterior median of the predictive distribution of y_j .

Errors are standardized by the standard deviation of the error and adjusted by the total error $\tau_j = \sqrt{s_j^2 + \sigma_j^2}$ as follows:

$$f_j = \frac{e_j}{sd(e_j)}$$

$$g_j = \frac{e_j}{\tau_j}$$

Coverage is given by $1/m \cdot \sum 1[y_j \geq l_j] \cdot 1[y_j \leq u_j]$ where l_j and u_j are lower and upper bounds of the 95% prediction interval and m denotes the number of left-out observations.

CHAPTER 3

RESULTS

3.1 Country estimates and projections

3.1.1 Illustrative country estimates and projections

Selected countries illustrate the variability in age patterns, levels, and trends (see figure 3.1.1). Swaziland appears to have a substantial increase in spacing among adolescents 15-19, which almost matches the increase seen in age groups 20-24 and 25-29. Ghana appears to limited growth in adolescents while having tremendous growth in other age groups. Older age groups in both countries, 40-49, have limited growth and low levels of spacing.

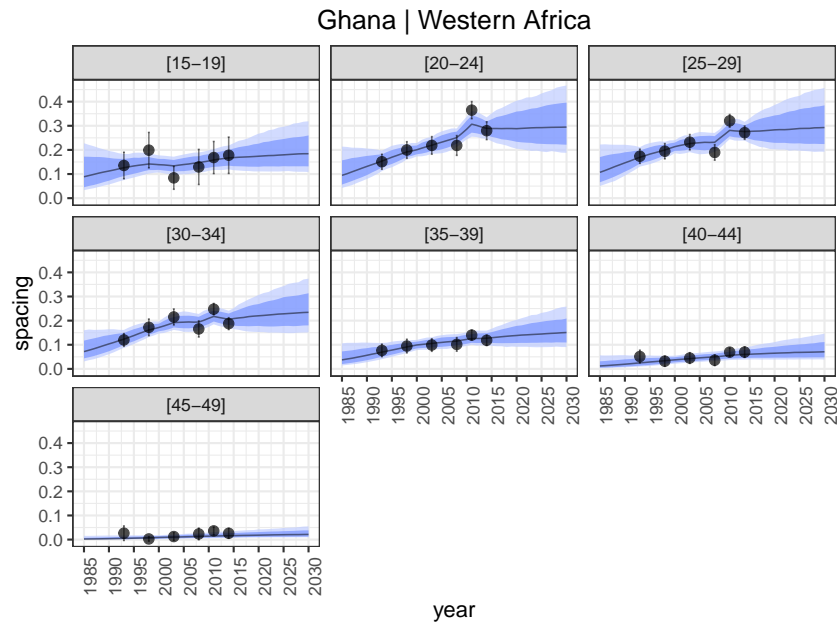
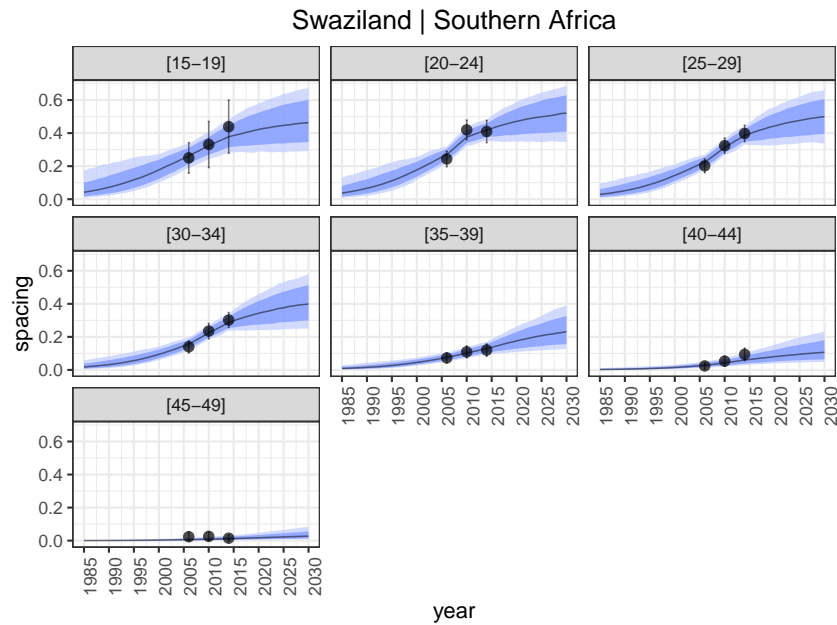
3.1.2 Choropleth maps

Estimates of spacing among ages 25-29 in 2018 are between 11% and 40% throughout most of sub-Saharan Africa. Southern and Eastern Africa have the highest estimated levels of prevalence in spacing. Across age groups we see the level of spacing peaks in ages 20-29 and decreases as age increases. (see figure 3.1.2)

Estimates of the change in spacing among 25-29 between 2000-2019 are below 10% across most of sub-Saharan Africa. There are a few countries that stand out in

Figure 3.1. Country estimates and projections 1985-2030 by 5-year age groups

The country projection plots show estimates of contraceptive use for spacing from 1985 through 2030 for each age group. Estimates represented by the solid line and 95% credible intervals shown in the light blue with 90% credible intervals shown in dark blue. Error bars around data points represent sampling variability.



Eastern and Southern Africa. Eswatini (Swaziland), Malawi, Rwanda, and Ethiopia see a change in prevalence greater than 20%.(see figure 3.1.2)

3.2 Model parameters

3.2.1 Asymptotes

Asymptote plots show estimates of country-age-specific and region-age-specific asymptotes where country asymptotes $\text{logit}(P_{a,c}) = \sum_{k=1}^K b_k(a)\alpha_{c,k}$ and regional asymptotes $\text{logit}(P_{a,r}) = \sum_{k=1}^K b_k(a)\alpha_{r,k}$.

The association between age and asymptote is similar across regions (see figure 3.2.1). Eastern Africa and Southern Africa have a comparatively high asymptote for adolescents. The greatest asymptote in these regions is for ages 20-24 compared to 25-29 in other regions. Western Africa and Middle Africa have lower asymptotes for all age groups. In these regions, the curve peaks at 25-29 and the differences among age groups are less pronounced.

Within regions, country asymptotes tend to have a similar trend across age groups with some variation (see figure 3.9). In Eastern Africa, where we have several countries, the similarities are evident. For most countries in this region, the asymptote is quite high in ages 15-19 and peaks in ages 20-24, with some variation in the general level. Plots for other regions are included in the Appendix (see figure A.1).

Figure 3.2. Estimates of spacing for the year 2019, ages 15-34

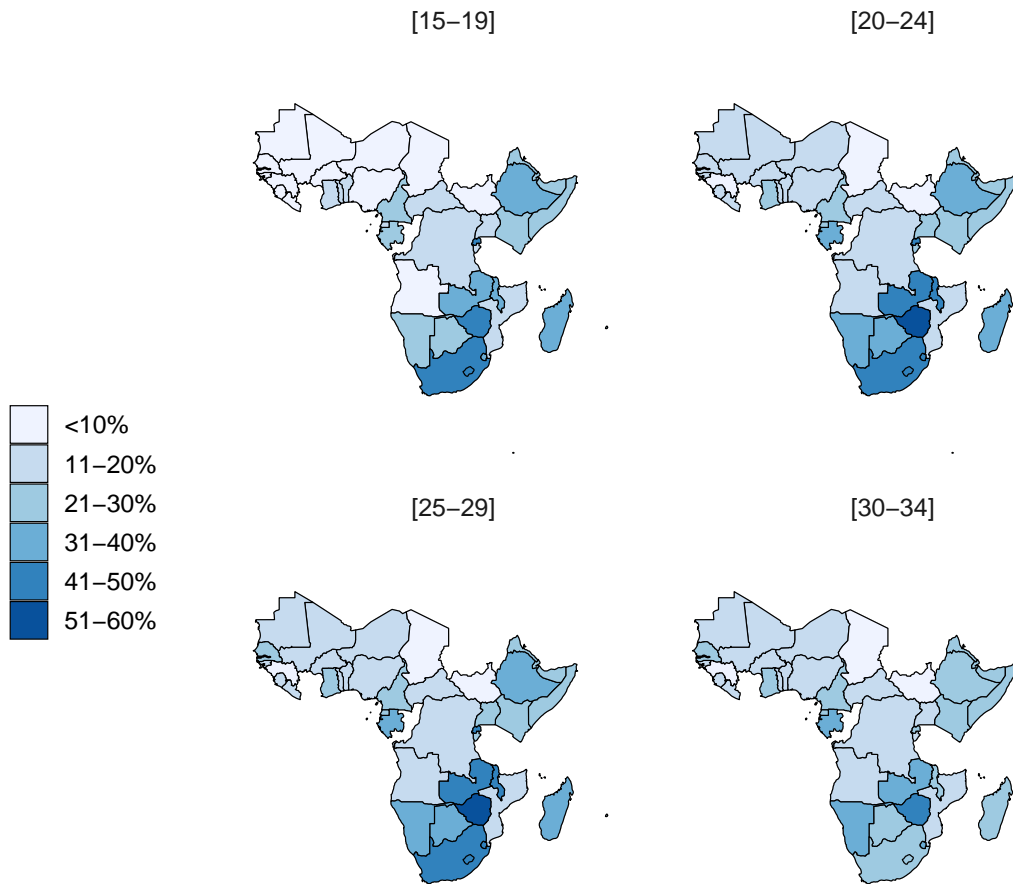


Figure 3.3. Estimates of spacing for the year 2019, ages 35-49

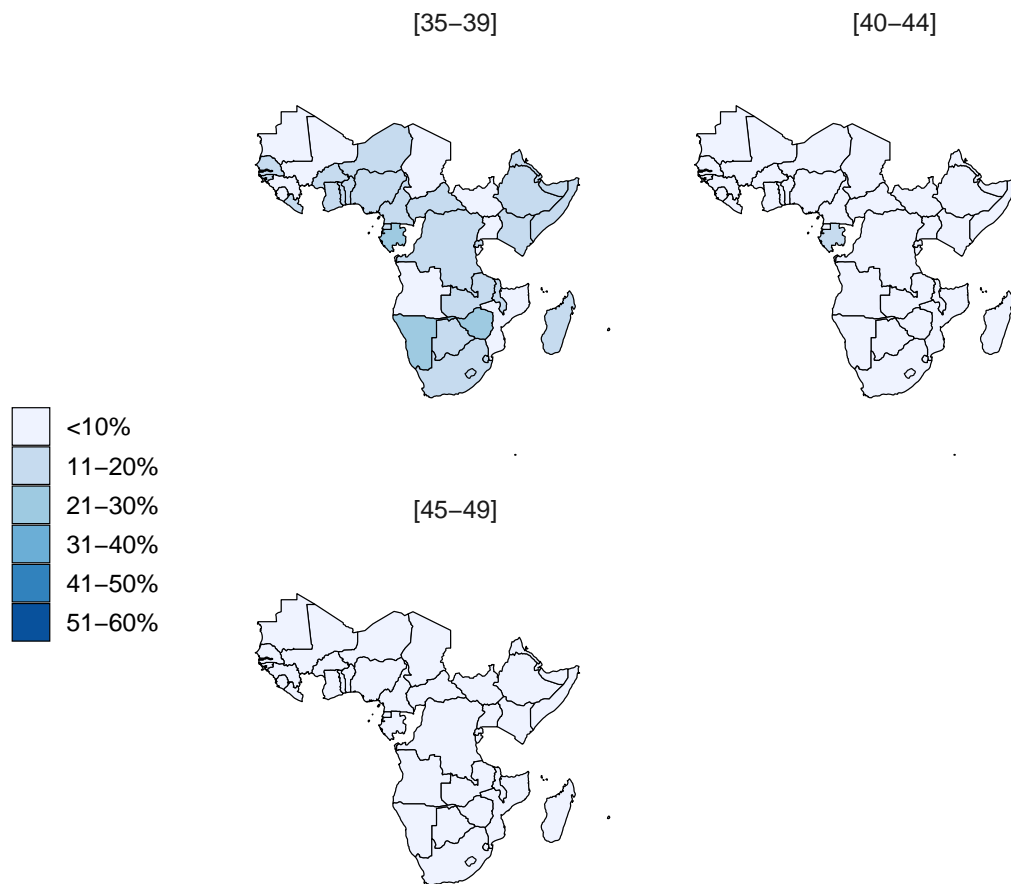


Figure 3.4. Estimate change in spacing (2000-2019), ages 15-34

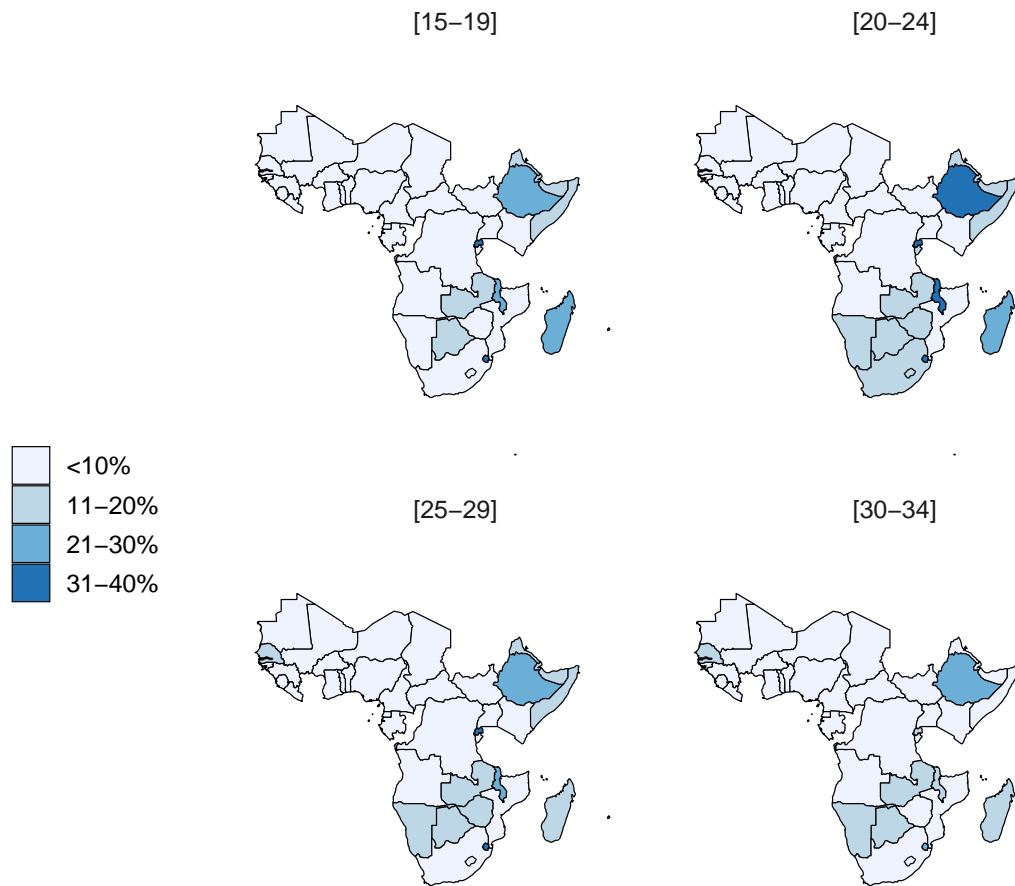


Figure 3.5. Estimate change in spacing (2000-2019), ages 35-49

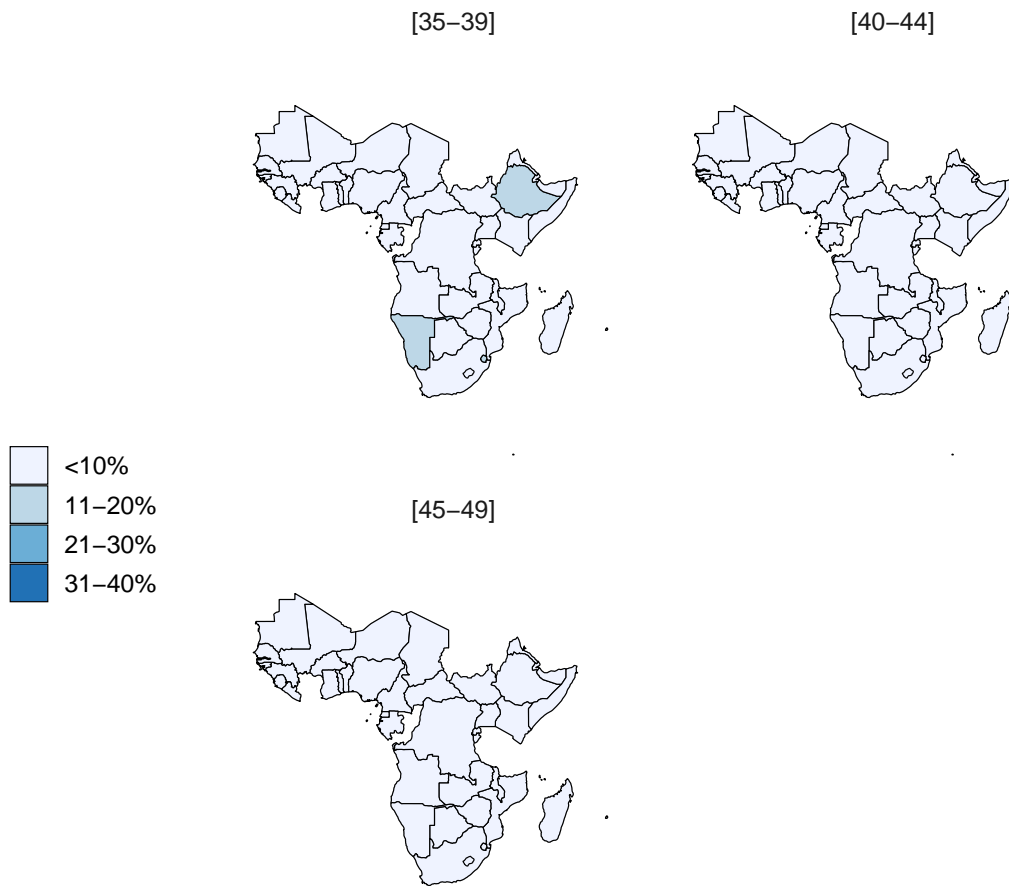
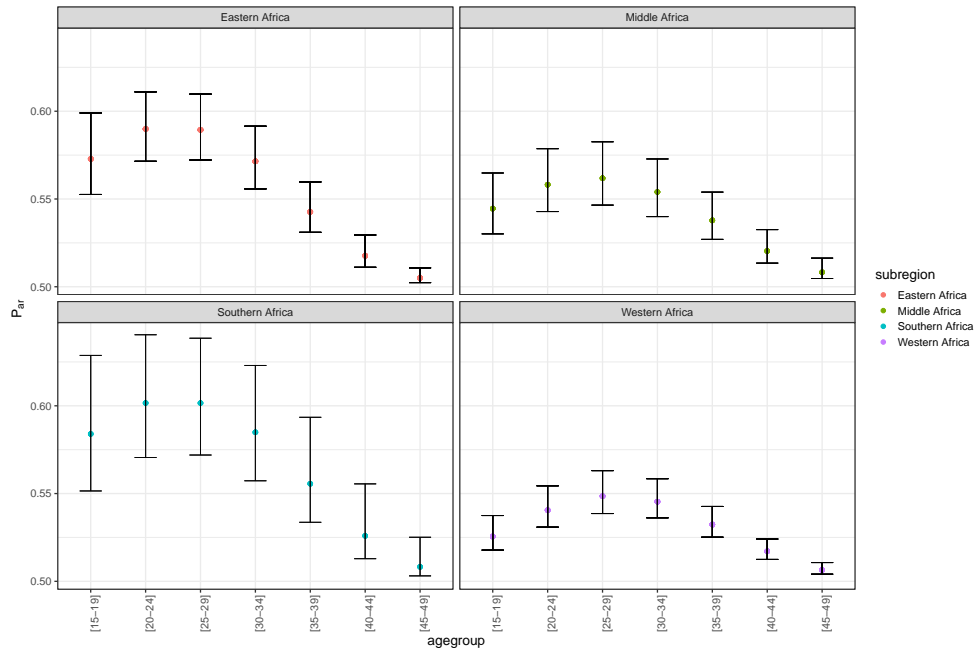
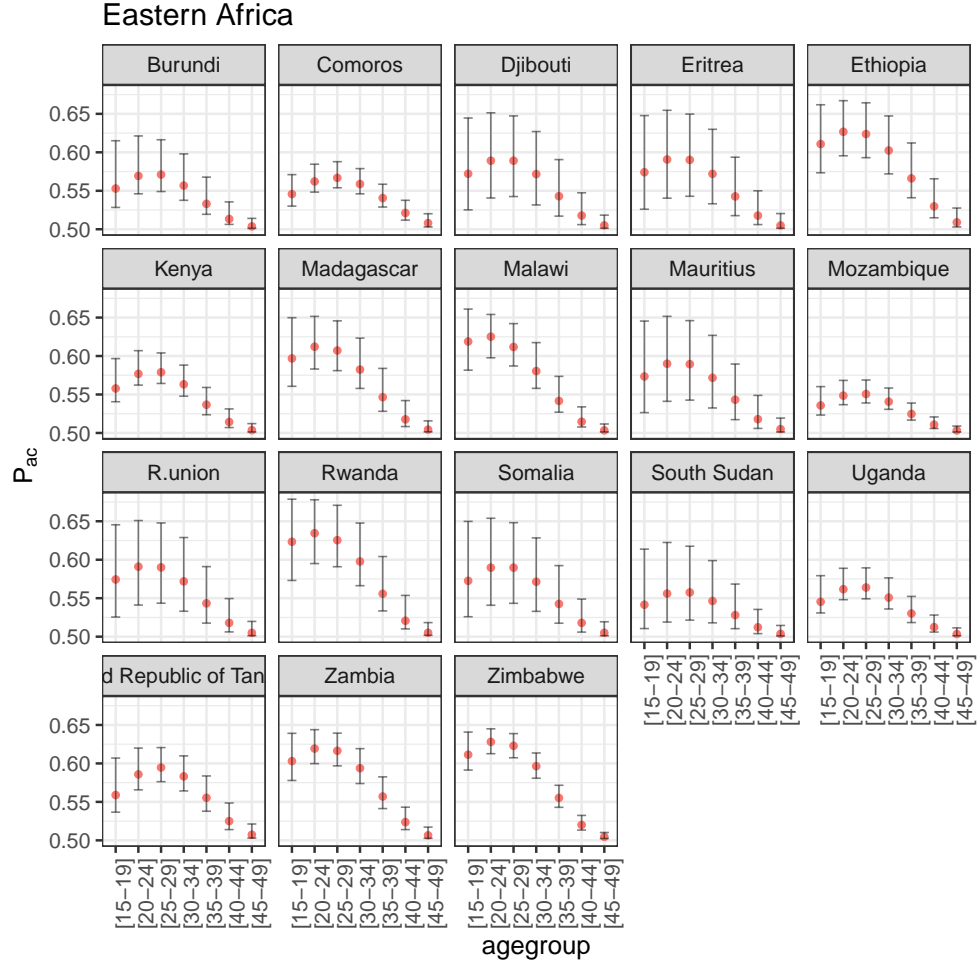


Figure 3.6. Regional asymptotes



Regional asymptotes $\text{logit}(P_{a,r}) = \sum_{k=1}^K b_k(a)\alpha_{r,k}$. Error bars around estimates represent the 95% credible intervals of the parameters.

Figure 3.7. Country asymptotes — Eastern Africa



3.2.2 Pace of contraceptive transition parameters

Pace of contraceptive transition plots show estimates of country-age-specific pace and region-age-specific pace $\omega_{a,c}$ and $\omega_{a,r}$, where $\omega_{a,c}$ is relative to an age specific asymptote as expressed below.

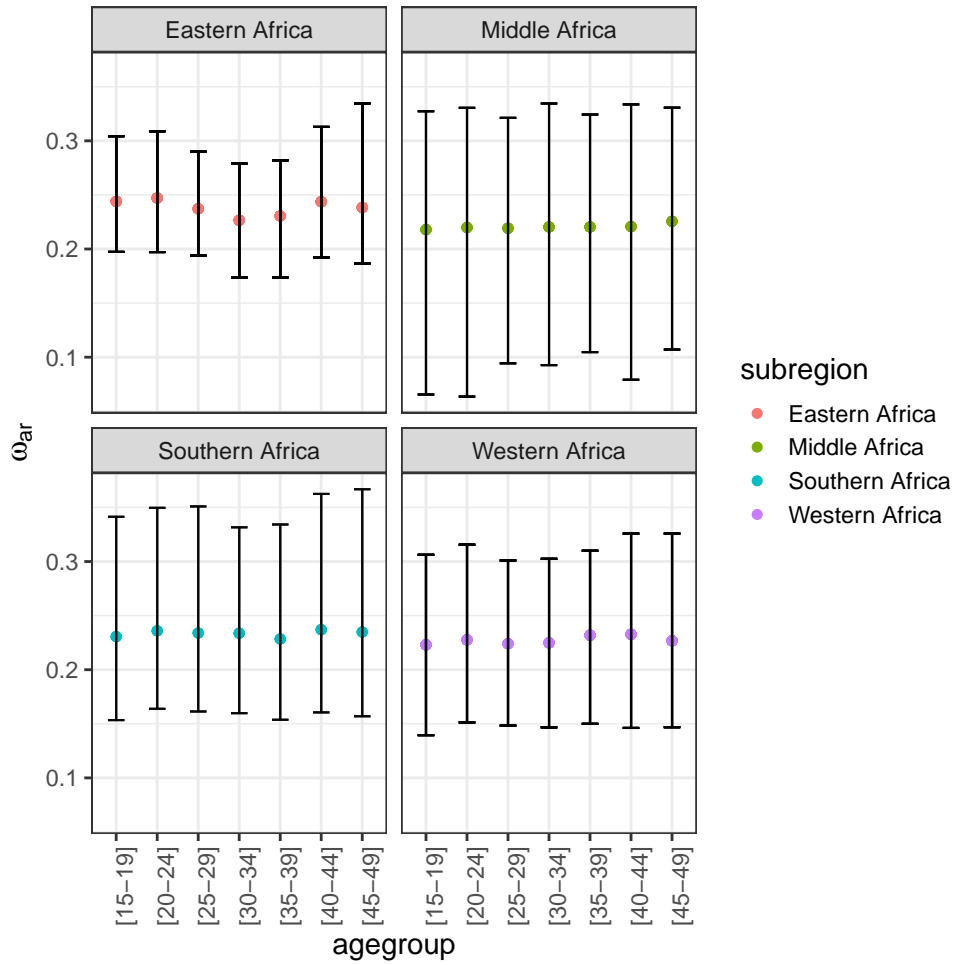
$$\frac{\eta_{a,c,t}^*}{P_{a,c}} = \frac{1}{1 + e^{-\omega_{a,c}(t-t_0)}}$$

Thus $\omega_{a,c}$ is the difference in scaled prevalence as shown.

$$\omega_{a,c} = \text{logit}\left(\frac{\eta_{a,c,t}^*}{P_{a,c}}\right) - \text{logit}\left(\frac{\eta_{a,c,t-1}^*}{P_{a,c}}\right)$$

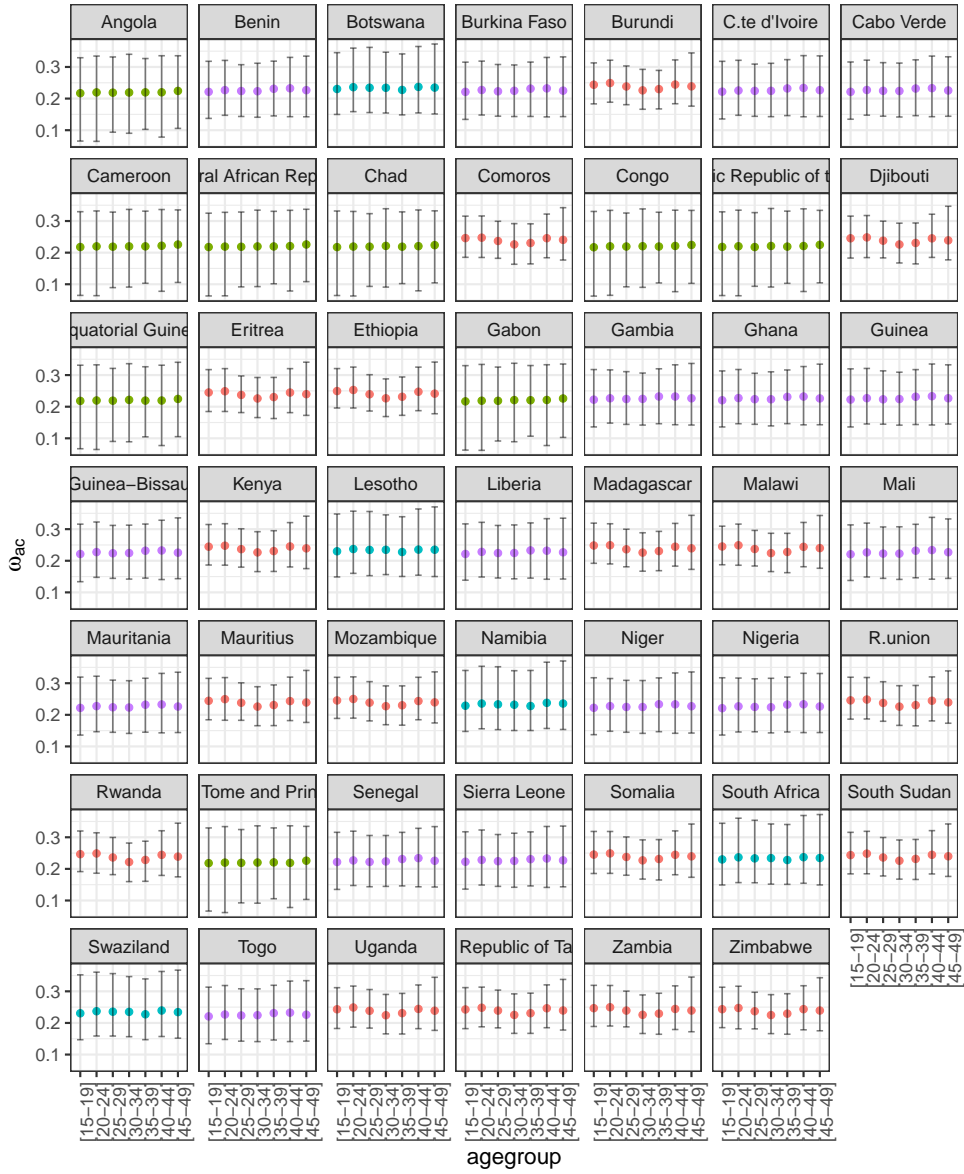
The pace is similar across age groups and across countries.

Figure 3.8. Regional pace parameters



Error bars
around estimates represent the 95% credible intervals of the parameters.

Figure 3.9. Country pace parameters



Error bars

around estimates represent the 95% credible intervals of the parameters.

Table 3.1. Validation results I: measures of absolute error

Out-of-sample	measure of center	e	f	g
1	median	.024	.445	.938
	mean	.034	.657	1.40
2	median	.009	.395	.378
	mean	.014	.649	.580

Table 3.2. Validation results II: measures of error

Out-of-sample	measure of center	e	f	g
1	median	.010	.194	.439
	mean	.011	.201	.358
2	median	.001	.085	.081
	mean	.004	.190	.143

3.3 Validation Results

The model performs well in validation. The coverage is around 95% for left out observations (see table 3.3). The mean and median errors are around 1% when leaving out recent data (Out-of-sample 1) and smaller when leaving out data at random (Out-of-sample 2) (see table 3.1). Overall, the model is well calibrated.

Table 3.3. Validation results III: coverage of left-out observations

Out-of-sample	Coverage
1	0.931
2	0.957

CHAPTER 4

DISCUSSION

Through the use of splines and Bayesian hierarchical models, we successfully estimate country-age specific prevalence of spacing for all countries in sub-Saharan Africa. Splines allow us to share information across age groups and Bayesian hierarchical models allow us to share information across countries.

A key feature of our model is the ability to detail age patterns. Our model structure reveals the diversity of progress across age groups while obtaining informed estimates for all ages.

Model estimates can be used for planning programming. For example, at the country level, analyzing historical trends by age can reveal stagnations or accelerations. Model projections can be used for target setting (Kantorová, New, Biddlecom, and Alkema 2017). The estimates can also be used for global monitoring and comparison, which can identify countries with a relatively slow or fast rate of change. Identified countries can then be used in case studies to compare and contrast family planning policies and programs.

Spacing appears to be gradually increasing over time with a more pronounced increase within ages 15-34 for select countries. The general trend across age groups is an increase in the prevalence of spacing from adolescents into adulthood, followed by

a sustained decrease. A possible explanation for this pattern is that women of ages 20-34 are more likely to be at parity 1+ and start spacing compared to adolescents. From here, as women age and families begin to reach preferred sizes, women may begin switching to limiting.

The current validation scheme is limited. Future validation schemes will leave out entire surveys to test the model's ability to provide estimates for countries without any data. Another limitation of the current set up is the calculation of sampling error. Future calculations of sampling error will account for the survey design.

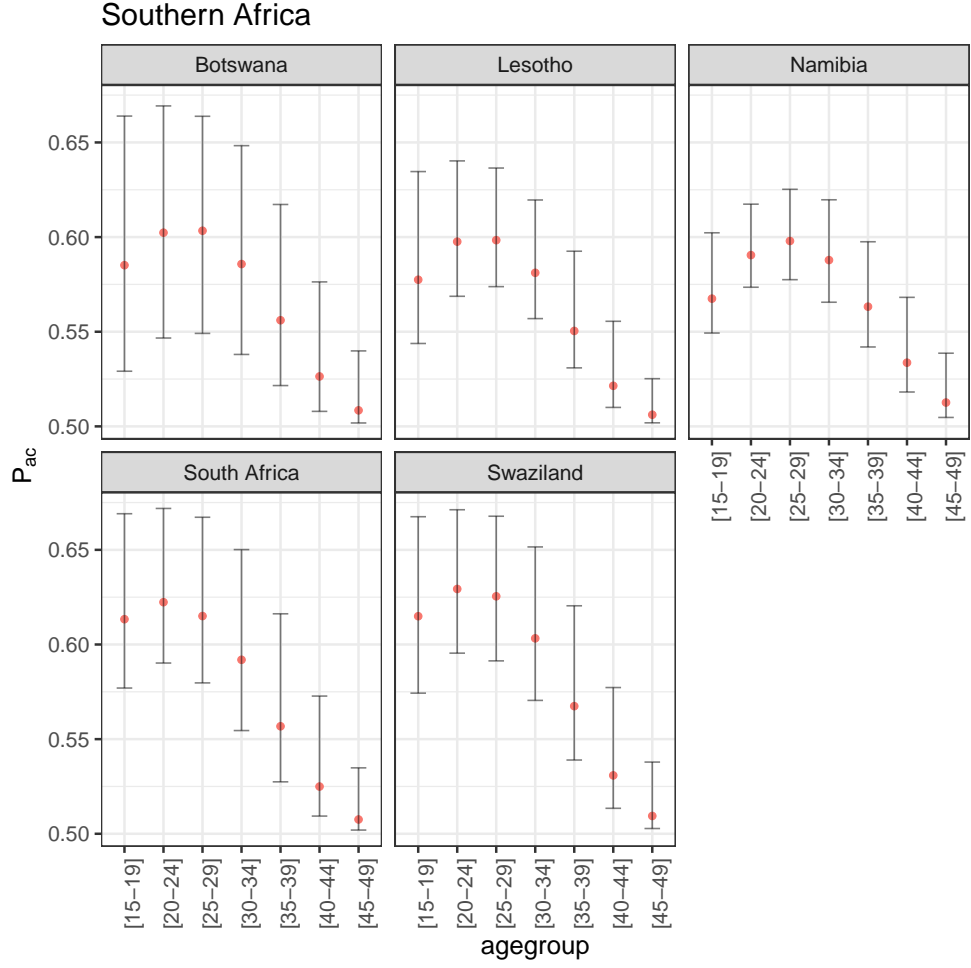
Possible extensions of the current model include the modeling of usage for limiting as well as unmet need for spacing and limiting. The estimation of these indicators for women who are not married or in a union can also be considered. Estimates of the complete model can be used for family planning programming or as inputs in other models. Other models include the estimation and projection of fertility, and the incidence of unintended pregnancies and abortions (Singh et al. 2010) (Bearak et al. 2018).

APPENDIX

SUPPLEMENTARY INFORMATION

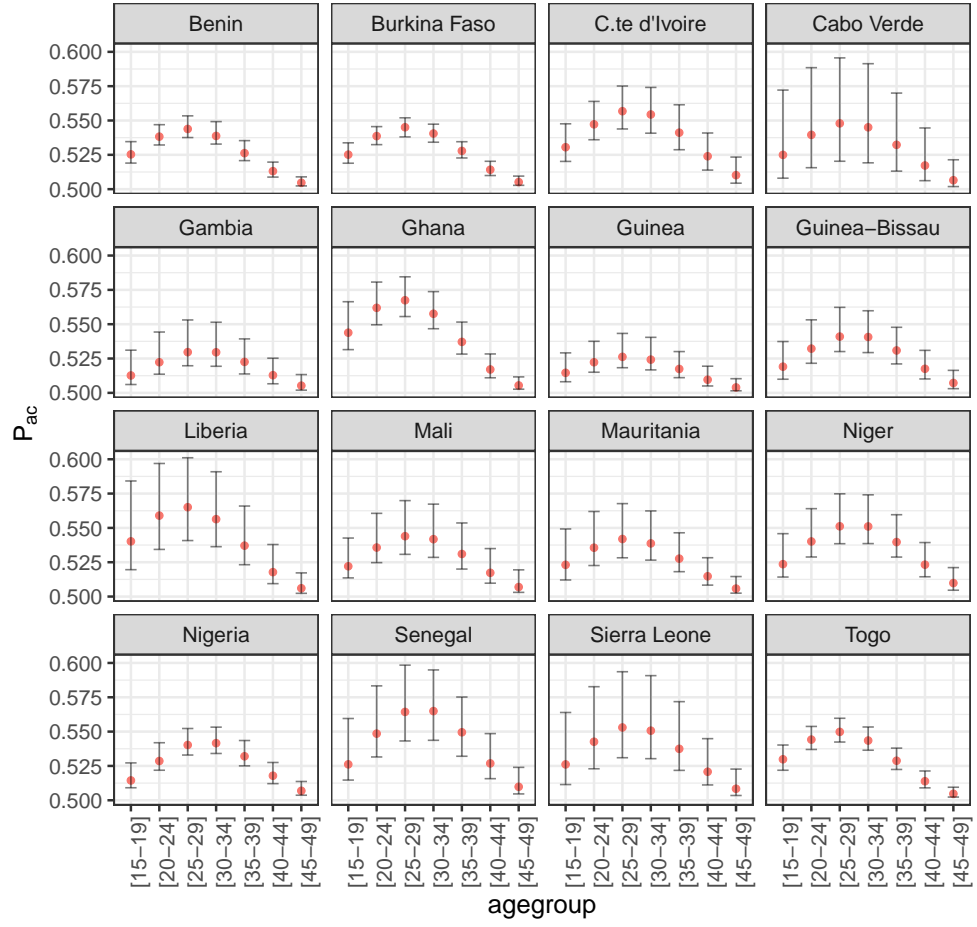
This appendix includes within-region asymptote plots which we left out of the main text. Also included are all country-age projections for 1985-2030.

Figure A.1. Country asymptotes by region



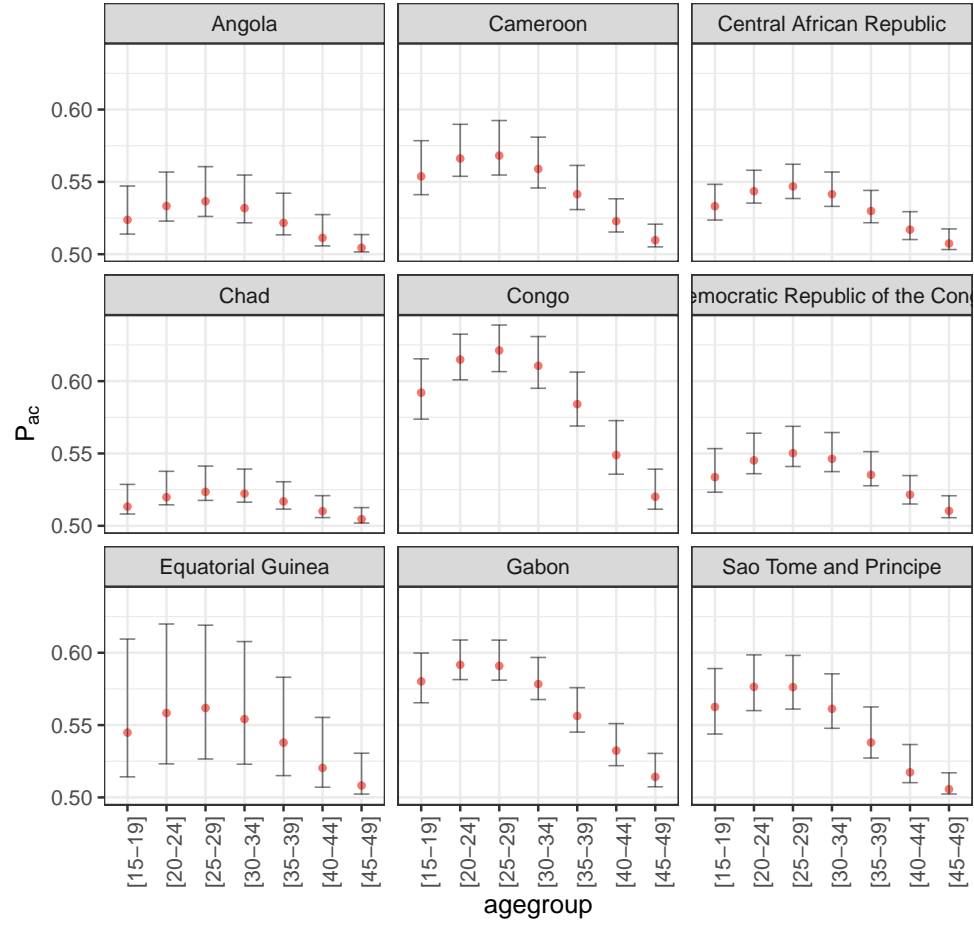
Country asymptotes $\text{logit}(P_{a,c}) = \sum_{k=1}^K b_k(a)\alpha_{c,k}$. Error bars around estimates represent the 95% credible intervals of the parameters.

Western Africa



Country asymptotes $\text{logit}(P_{a,c}) = \sum_{k=1}^K b_k(a)\alpha_{c,k}$. Error bars around estimates represent the 95% credible intervals of the parameters.

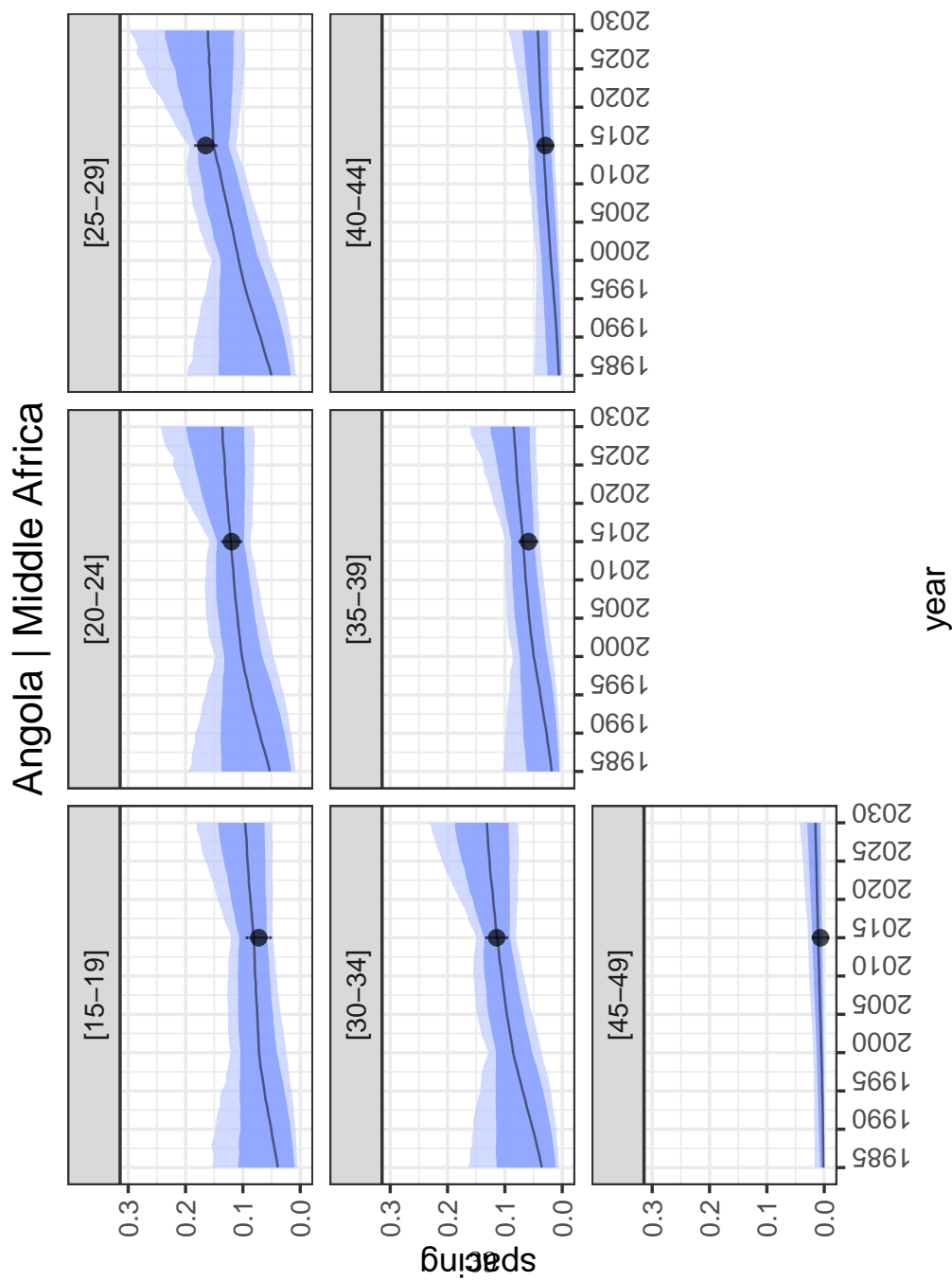
Middle Africa



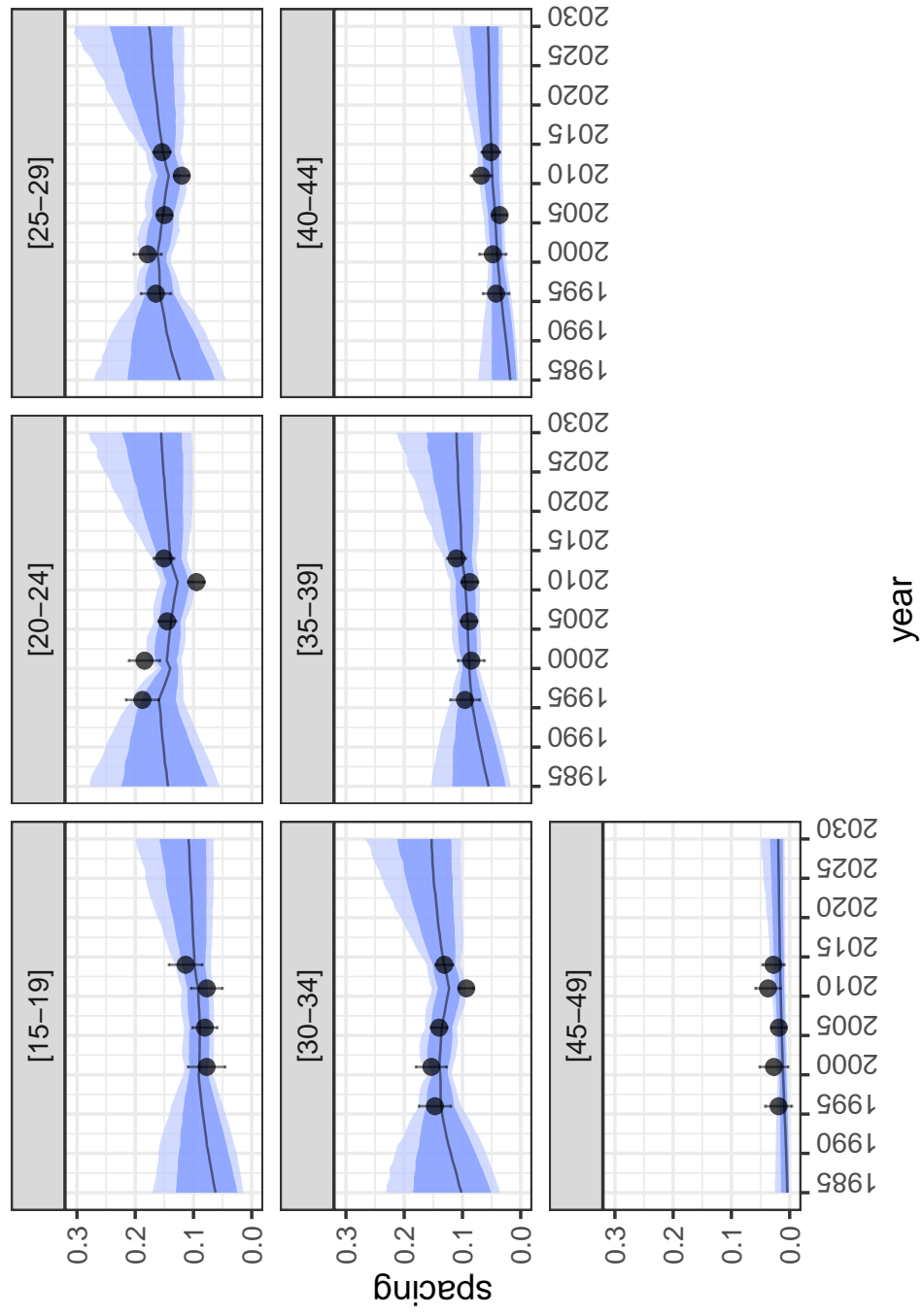
Country asymptotes $\text{logit}(P_{a,c}) = \sum_{k=1}^K b_k(a)\alpha_{c,k}$. Error bars around estimates represent the 95% credible intervals of the parameters.

Figure A.2. Country estimates and projections 1985-2030 by 5-year age groups

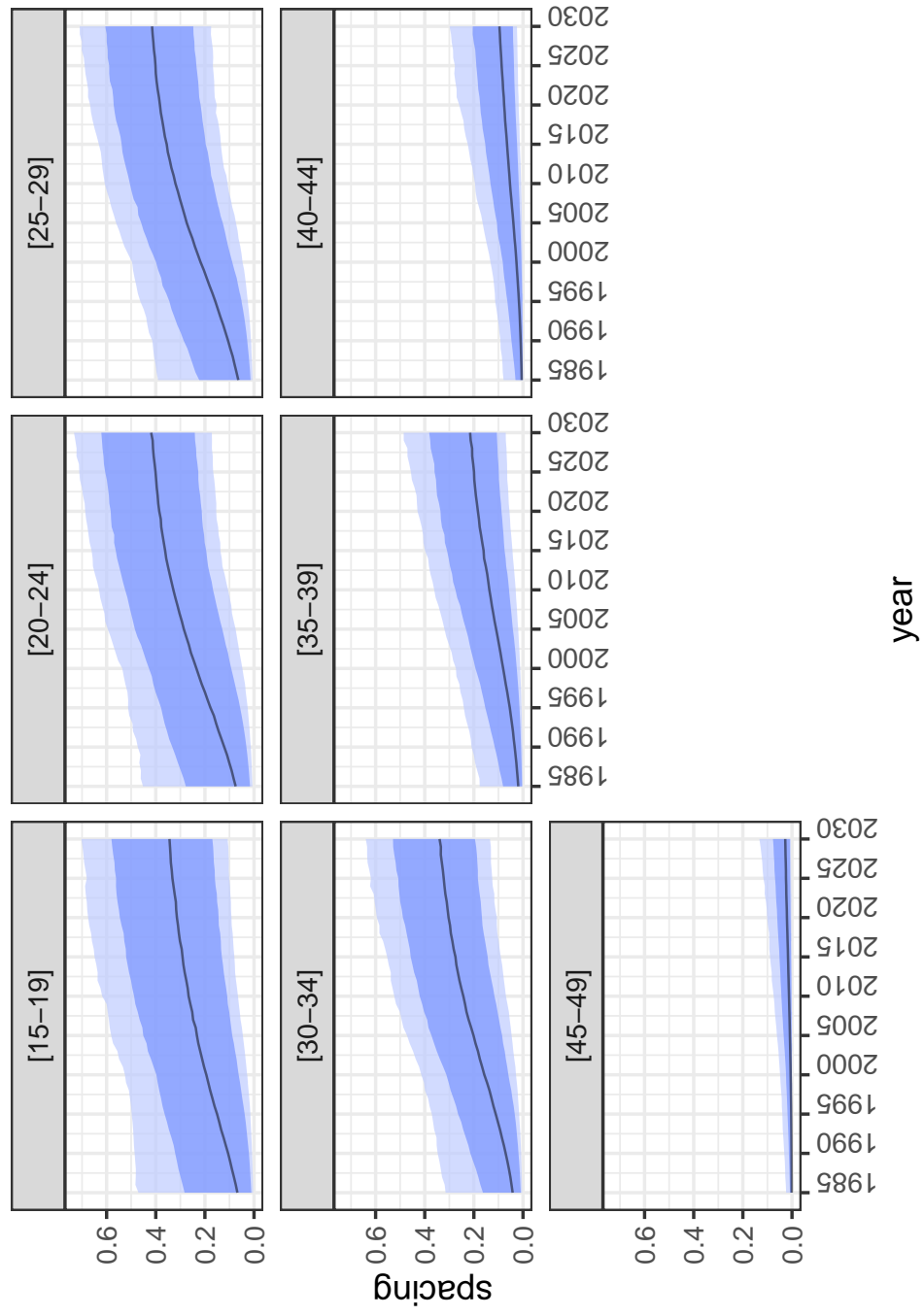
The country projection plots show estimates of contraceptive use for spacing from 1985 through 2030 for each age group. Estimates represented by the solid line and 95% credible intervals shown in the light blue with 90% credible intervals shown in dark blue. Error bars around data points represent sampling variability.



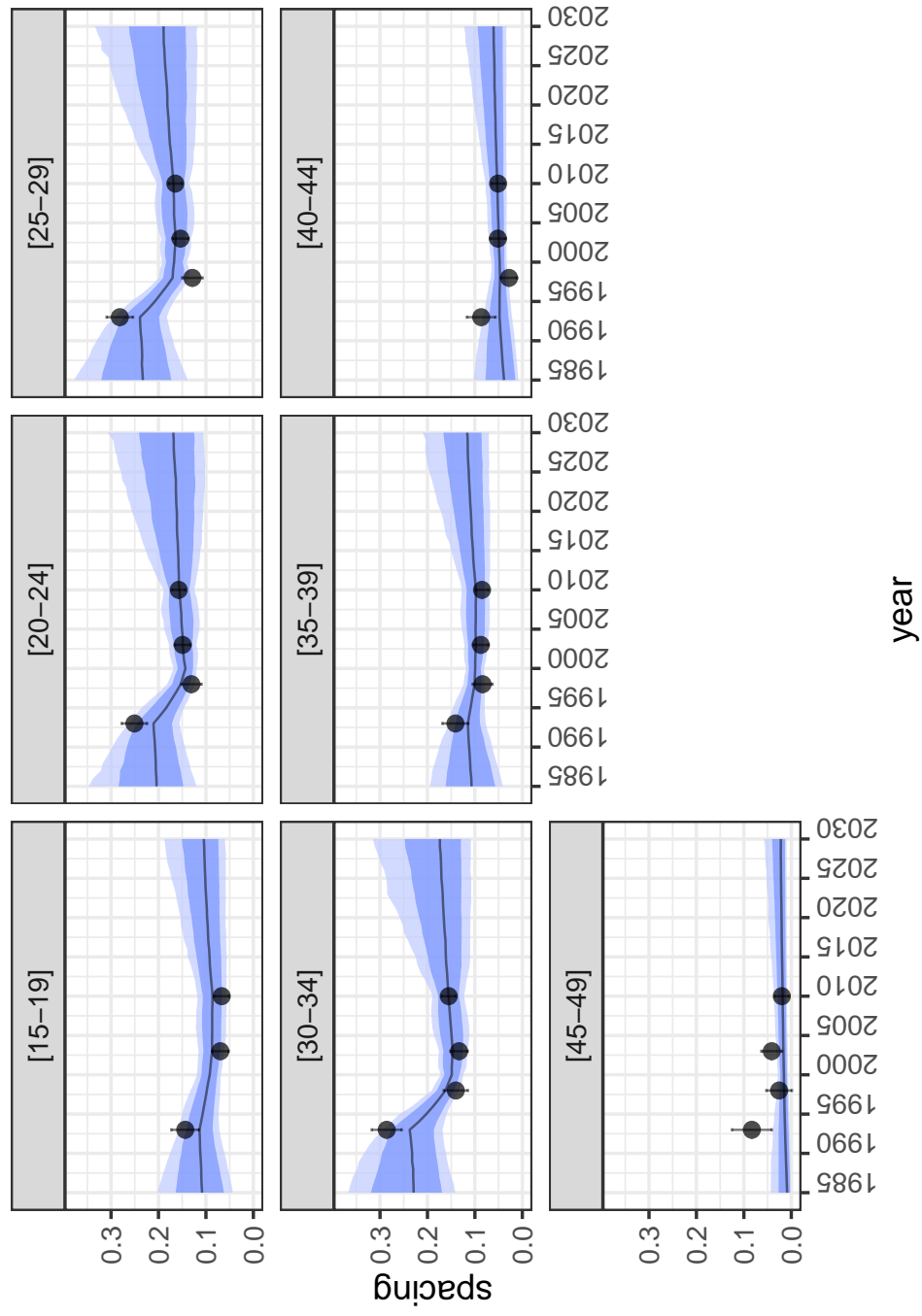
Benin | Western Africa



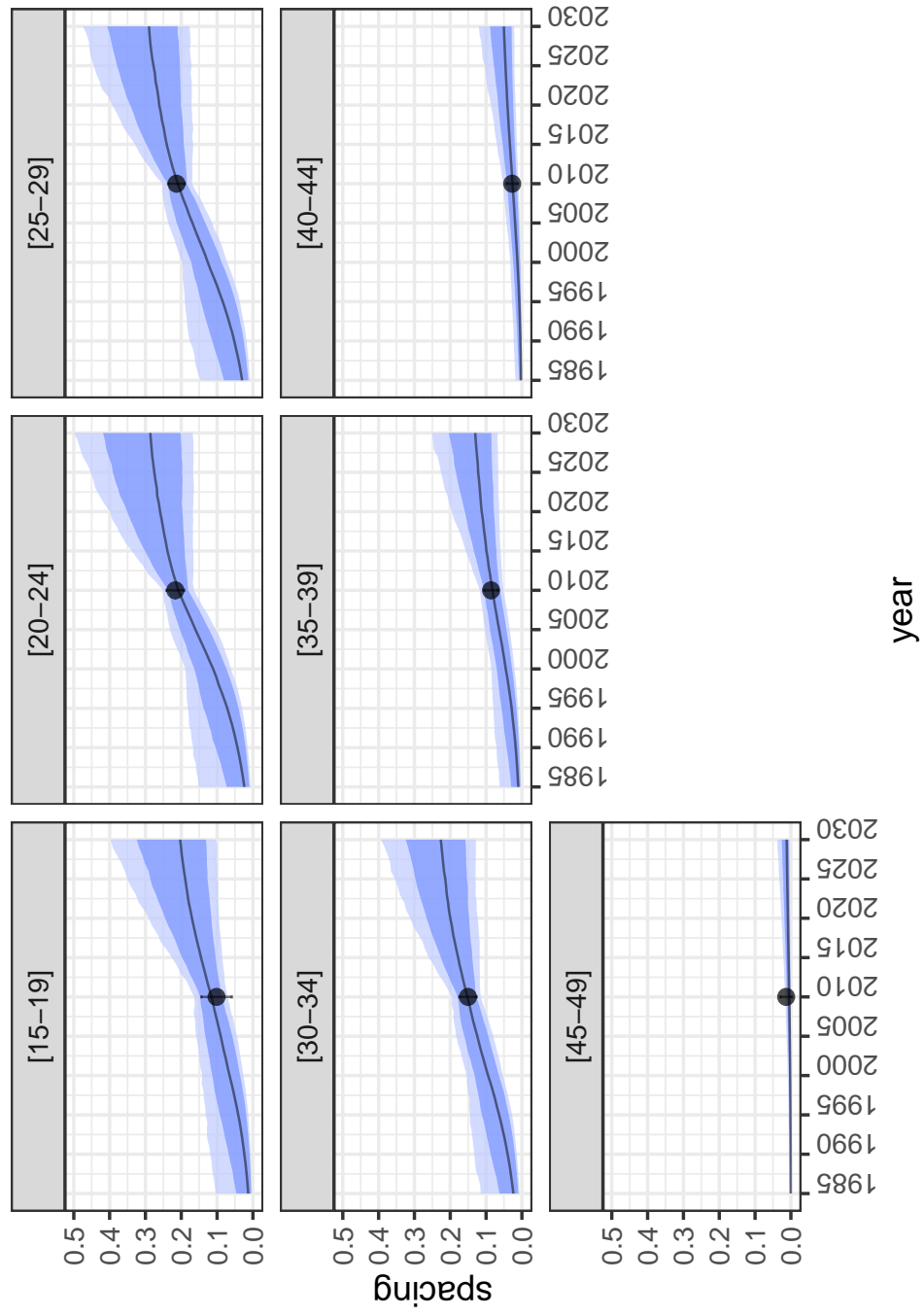
Botswana | Southern Africa



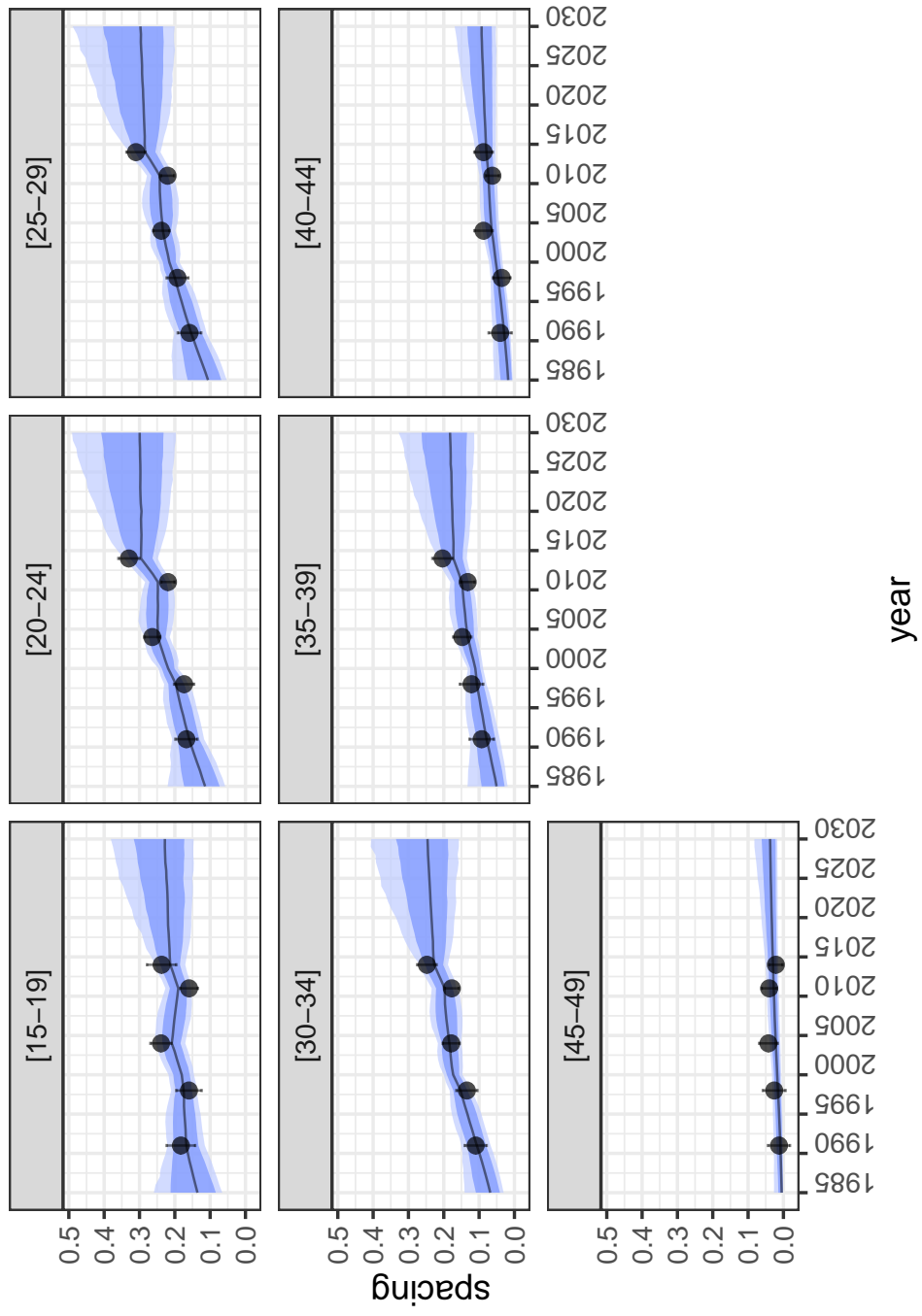
Burkina Faso | Western Africa



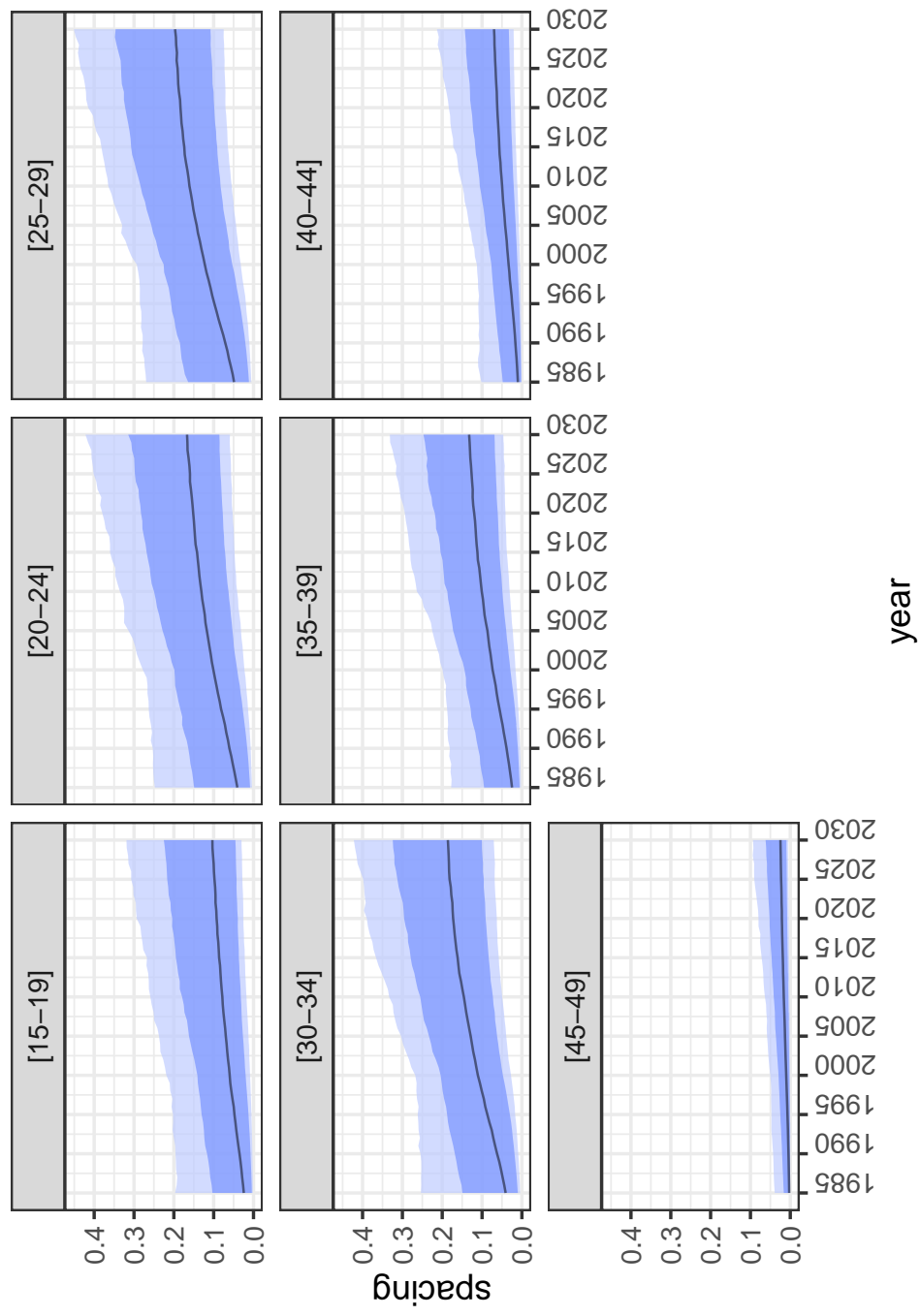
Burundi | Eastern Africa



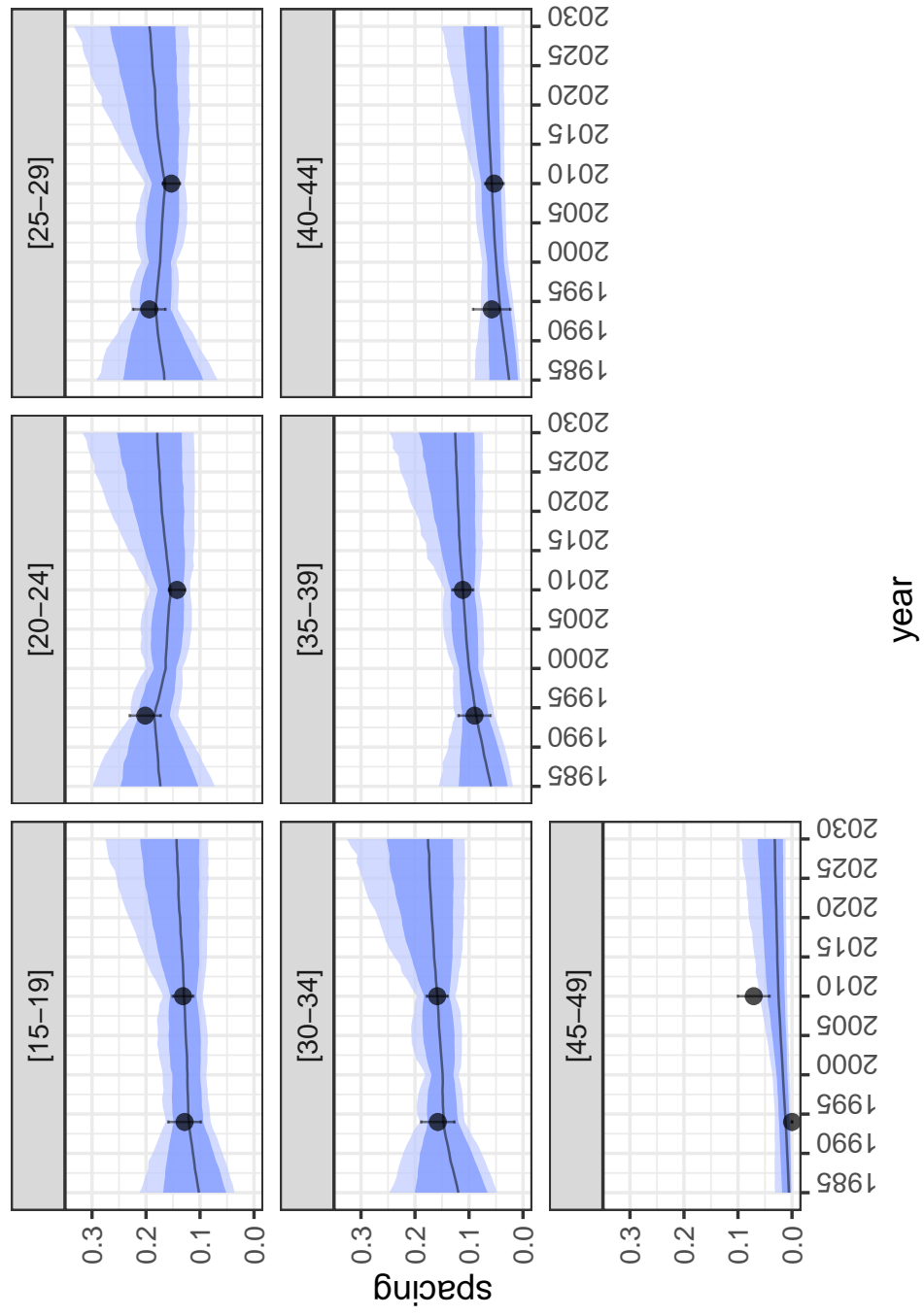
Cameroon | Middle Africa



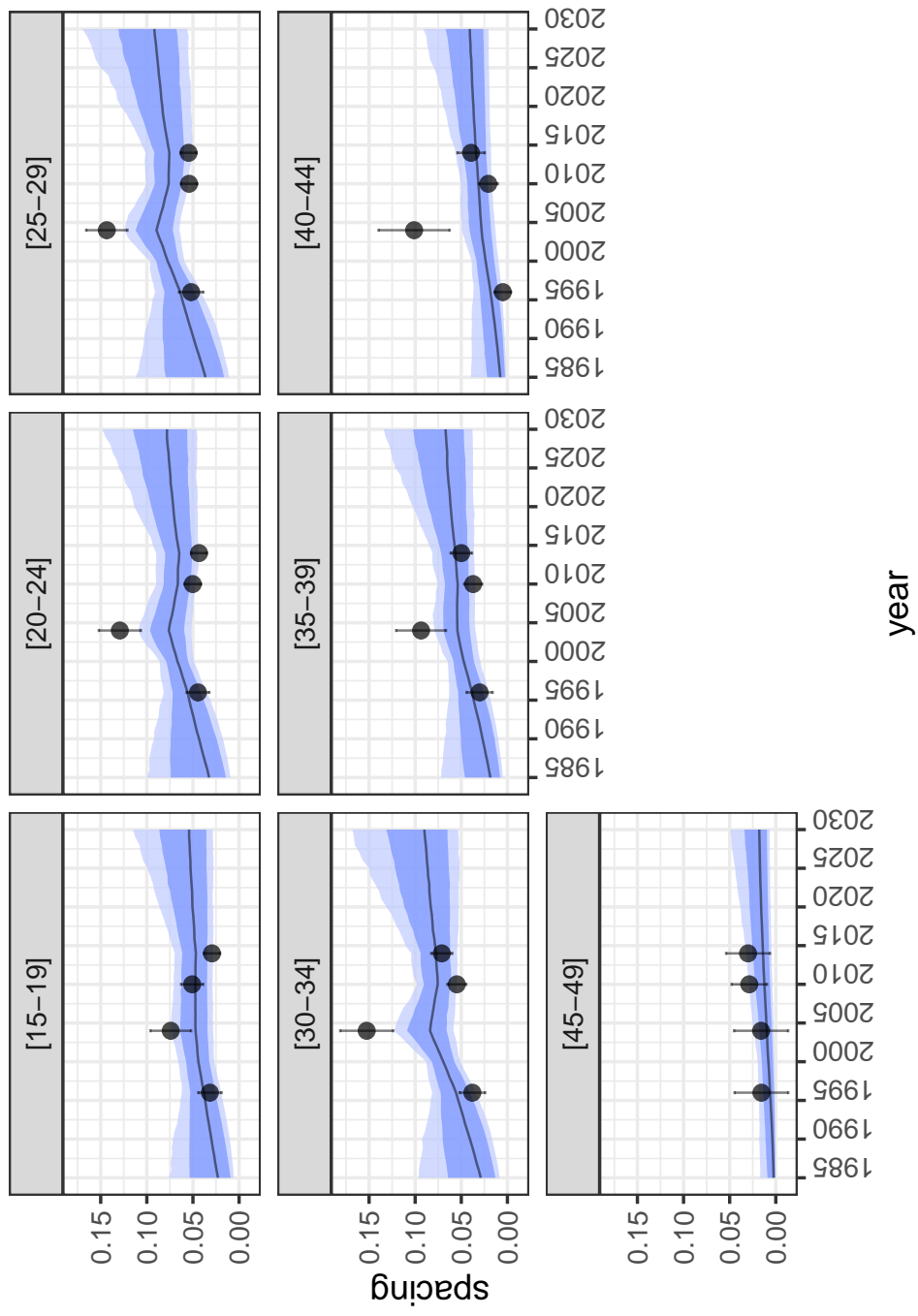
Cabo Verde | Western Africa



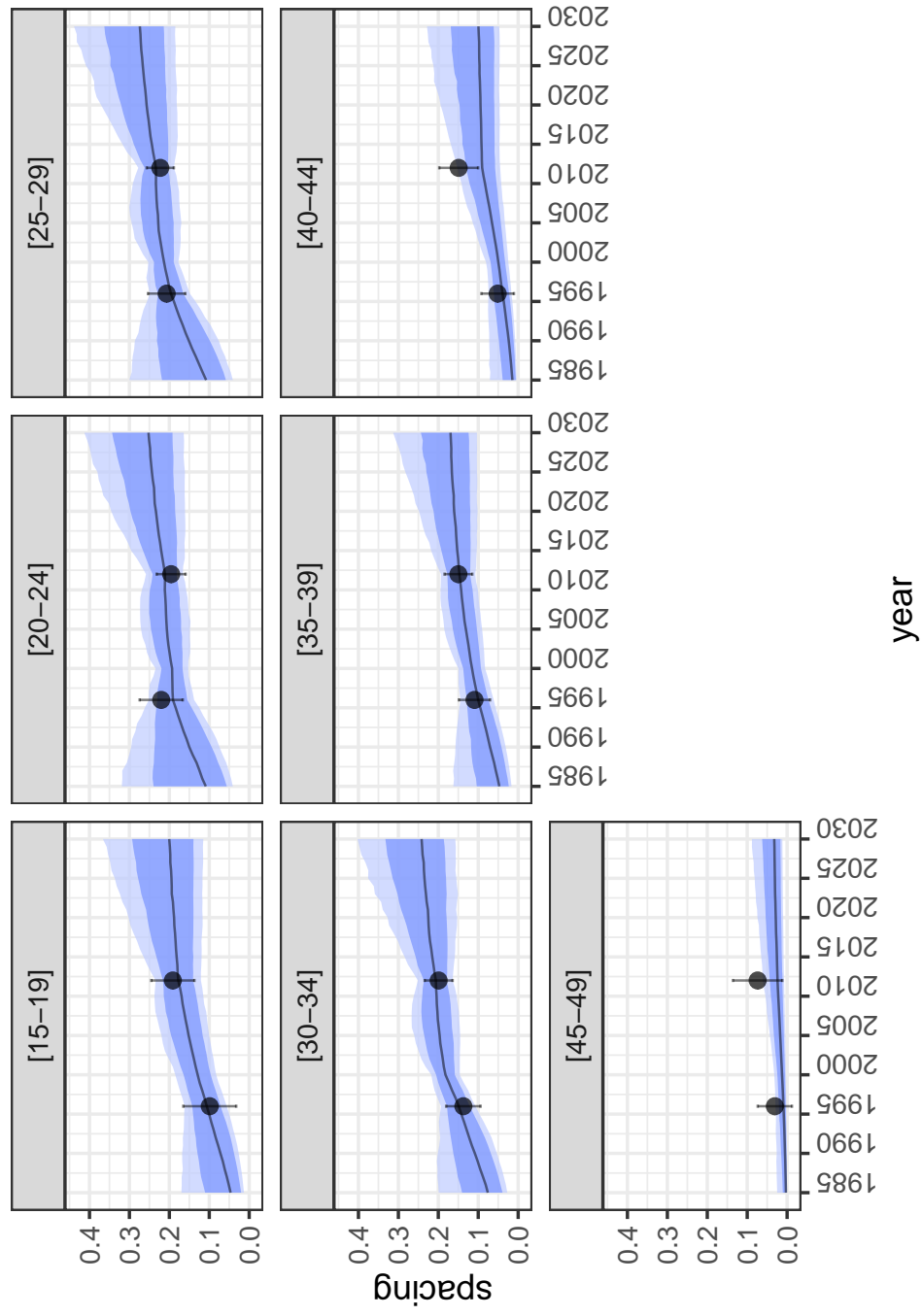
Central African Republic | Middle Africa



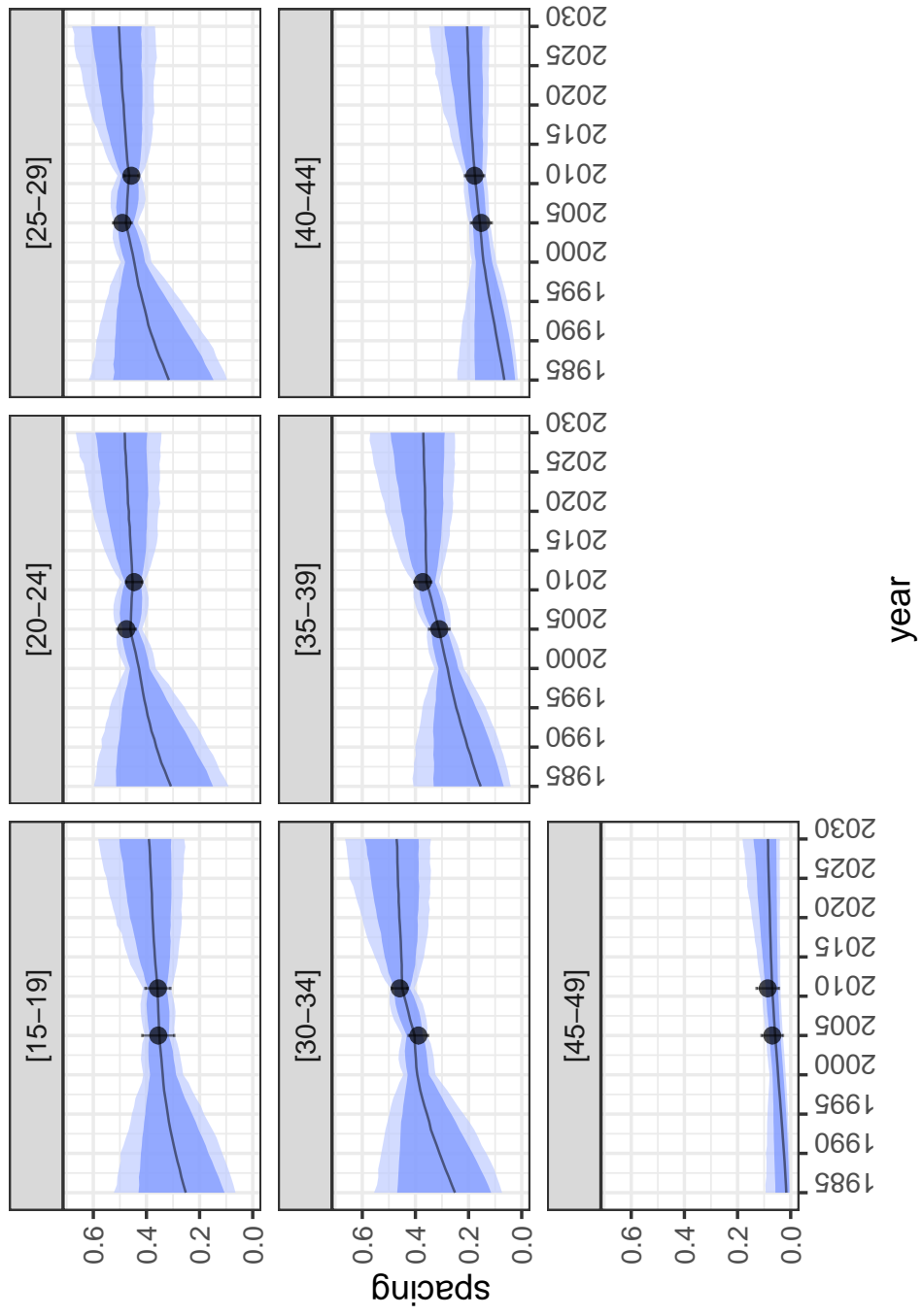
Chad | Middle Africa



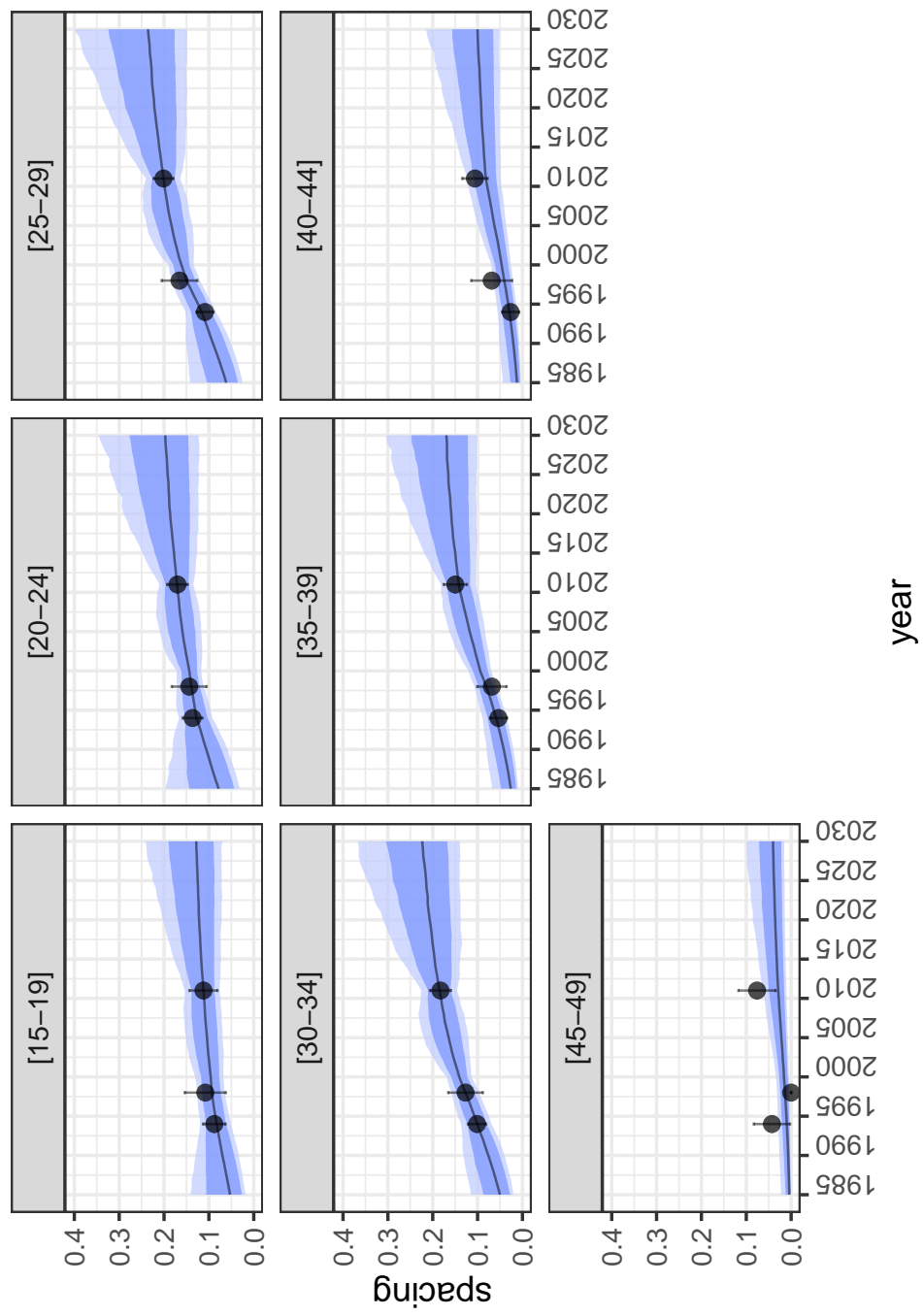
Comoros | Eastern Africa



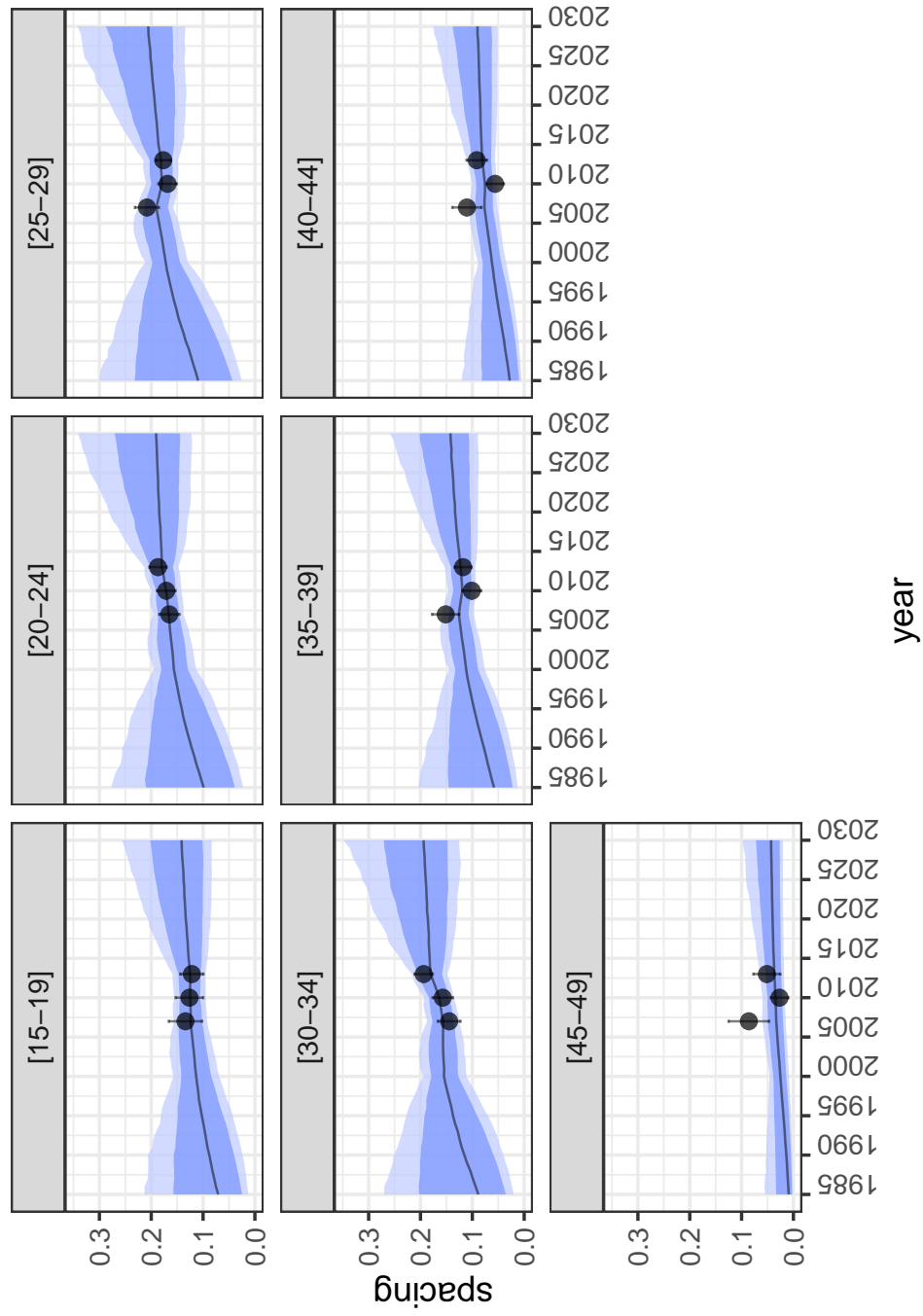
Congo | Middle Africa



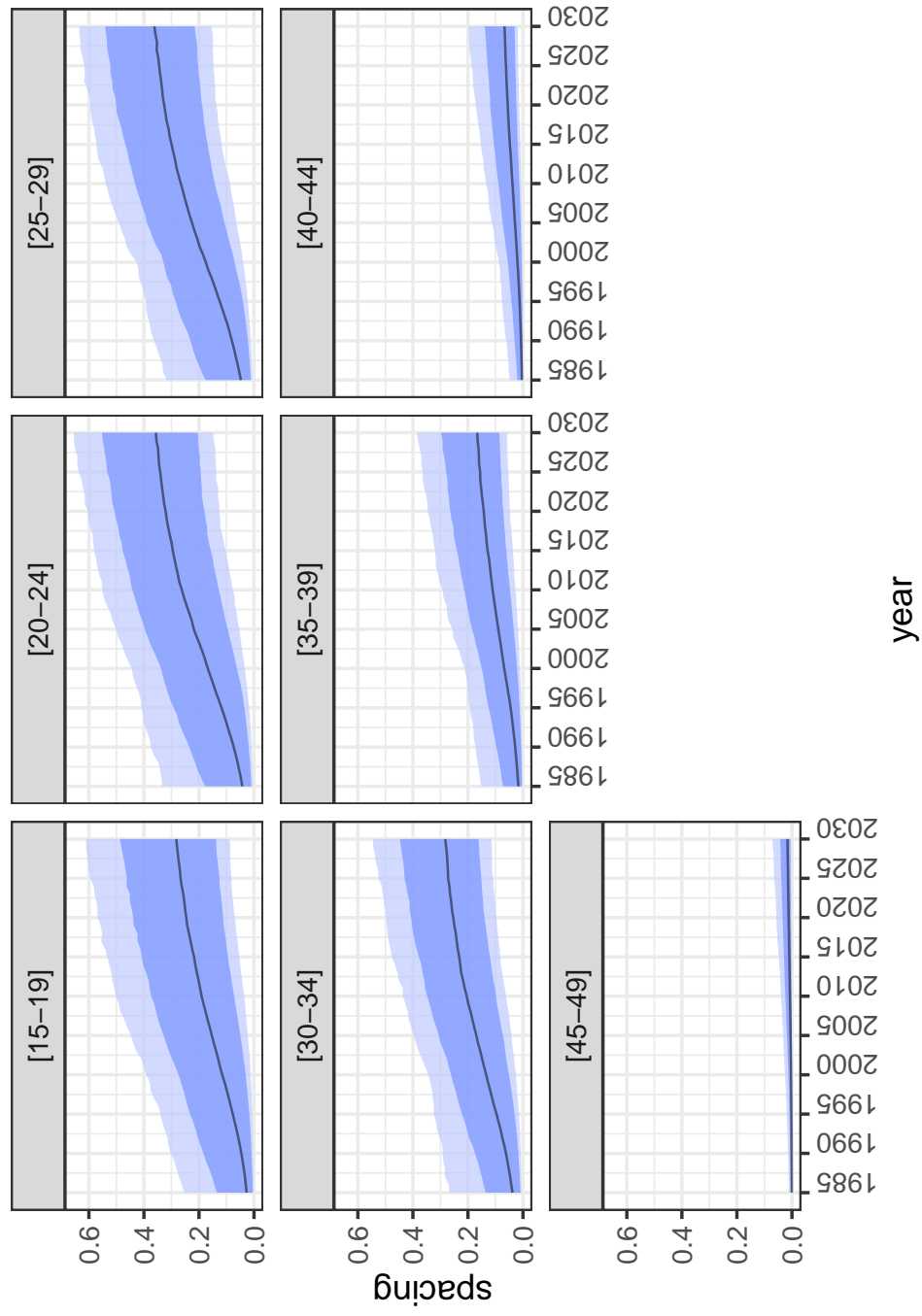
C.te d'Ivoire | Western Africa



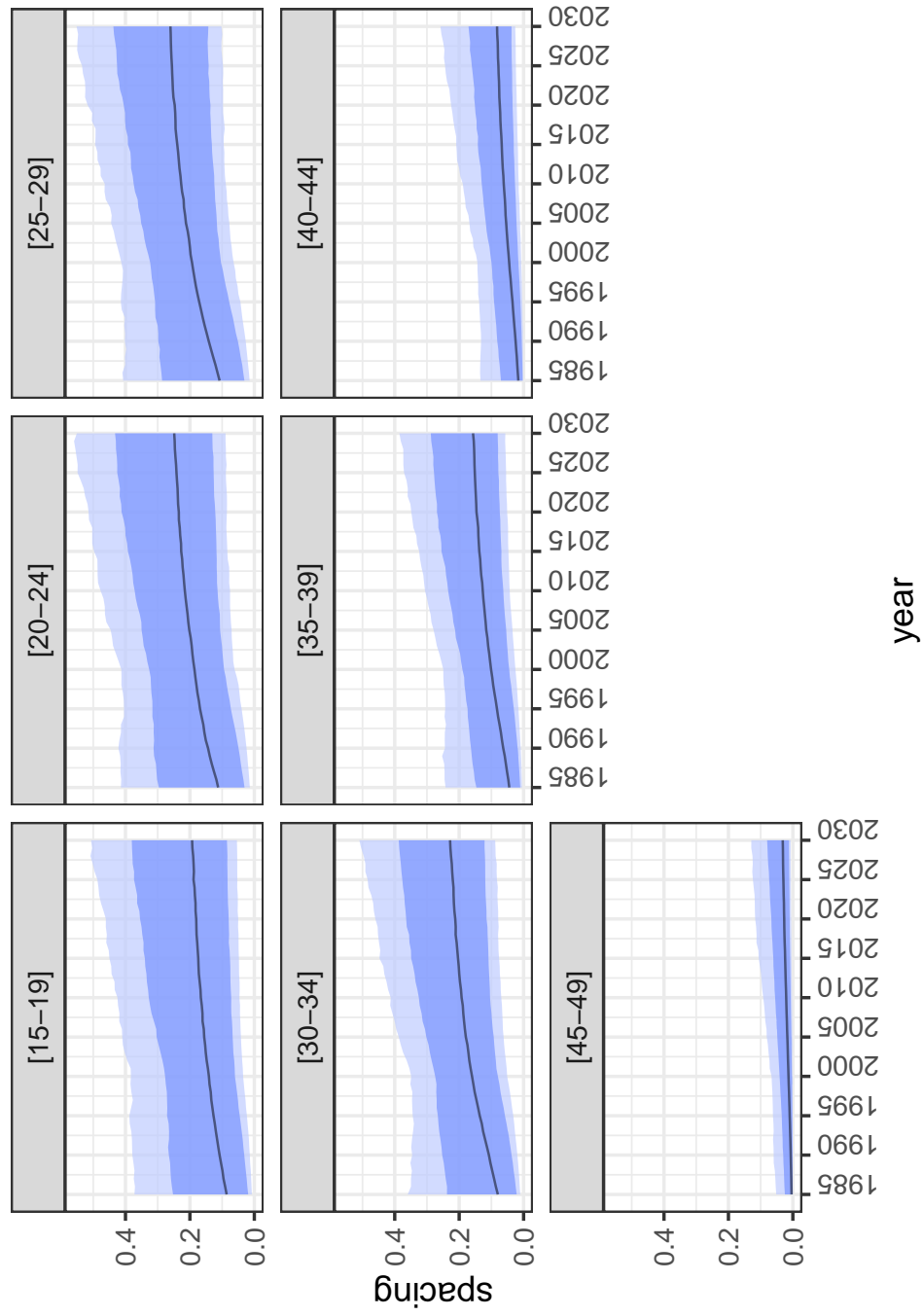
Democratic Republic of the Congo | Middle Africa



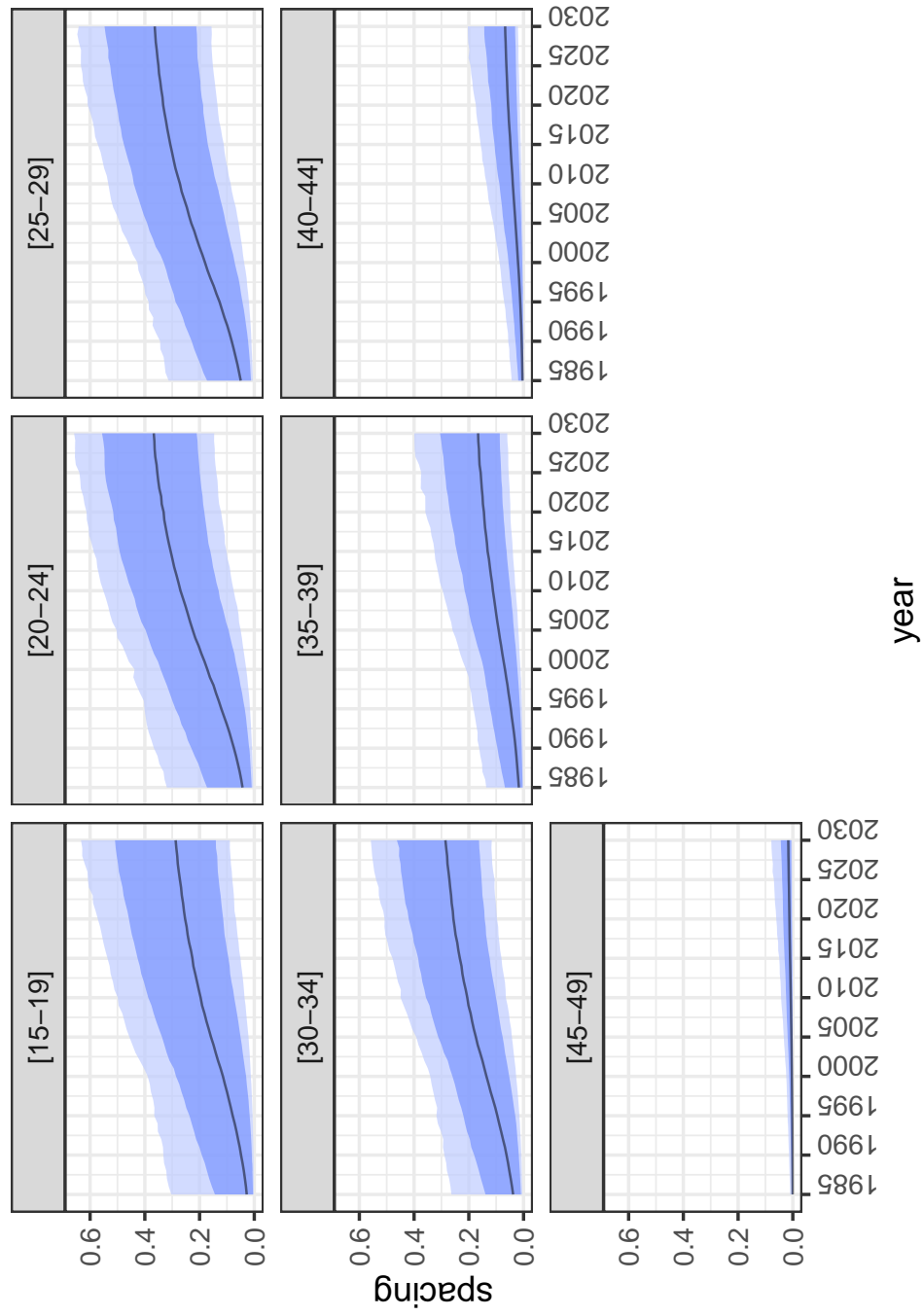
Djibouti | Eastern Africa



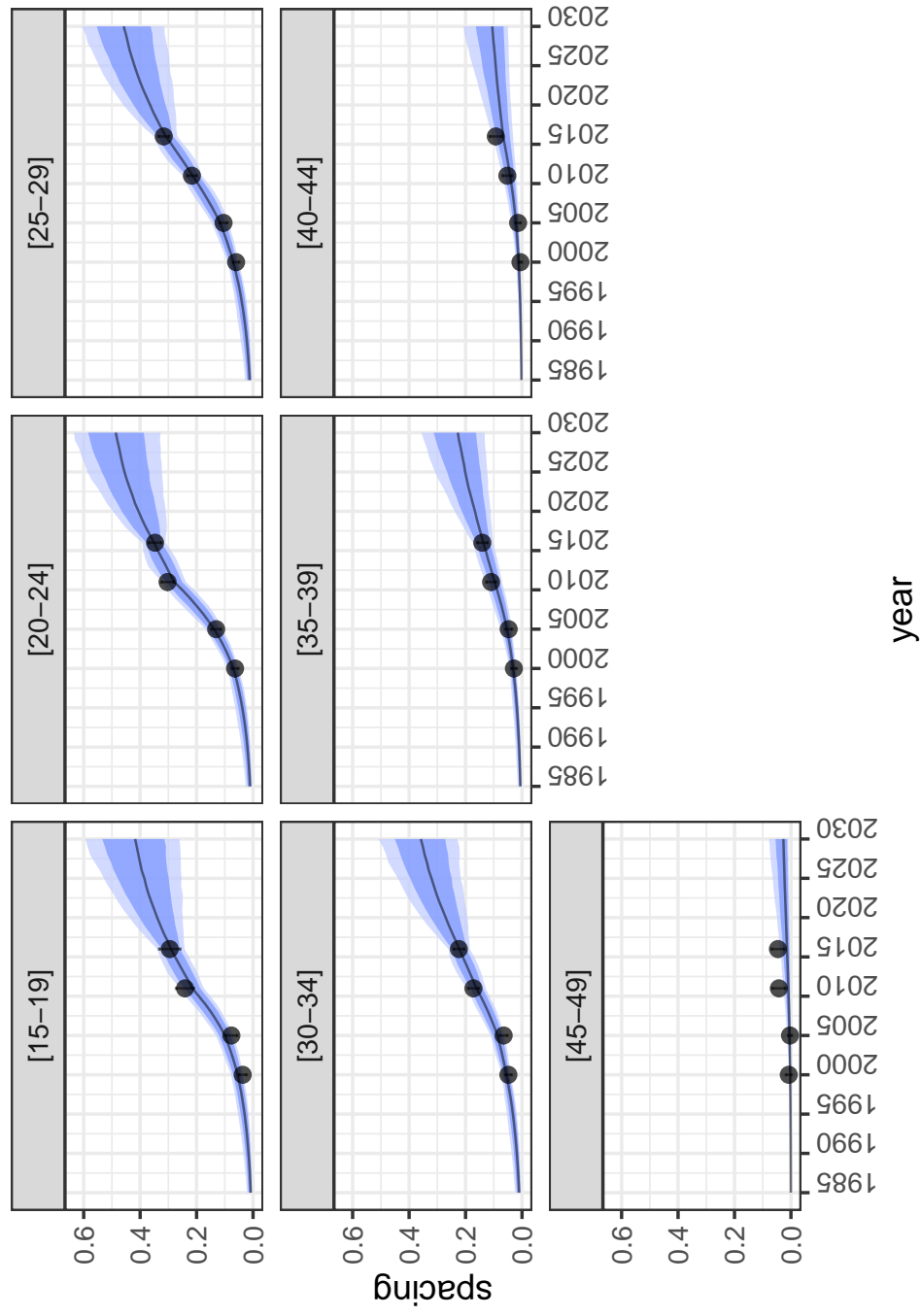
Equatorial Guinea | Middle Africa



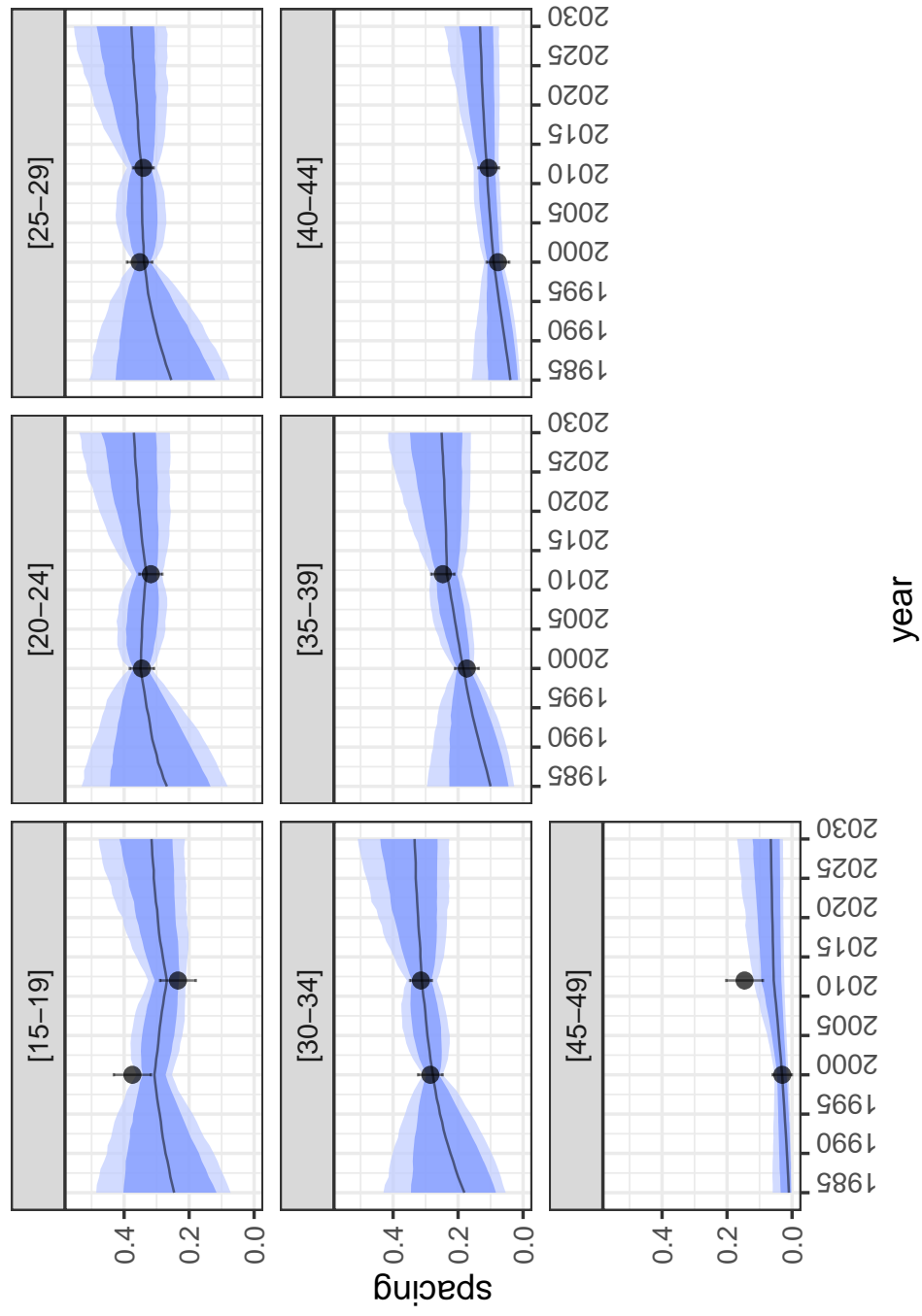
Eritrea | Eastern Africa



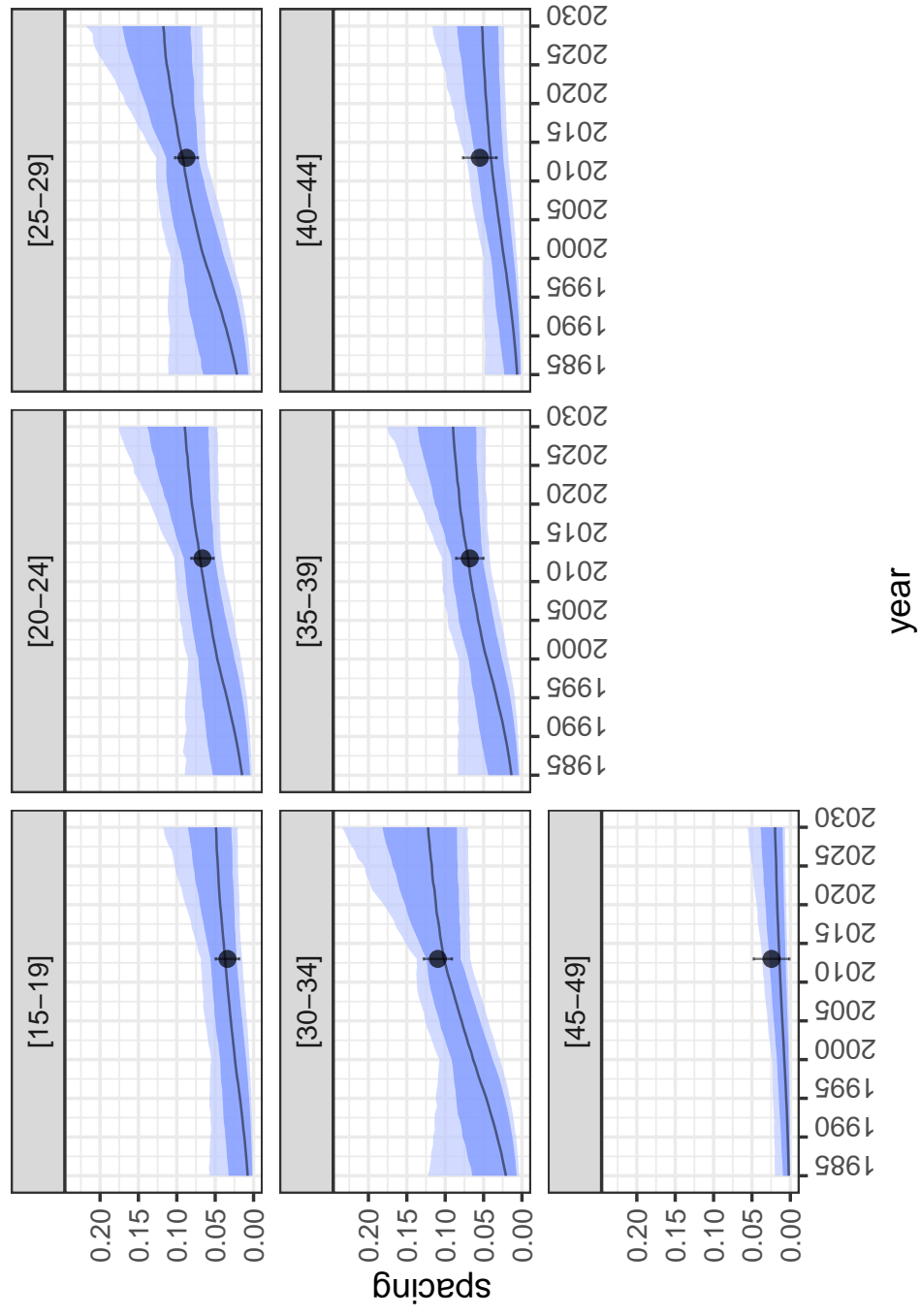
Ethiopia | Eastern Africa



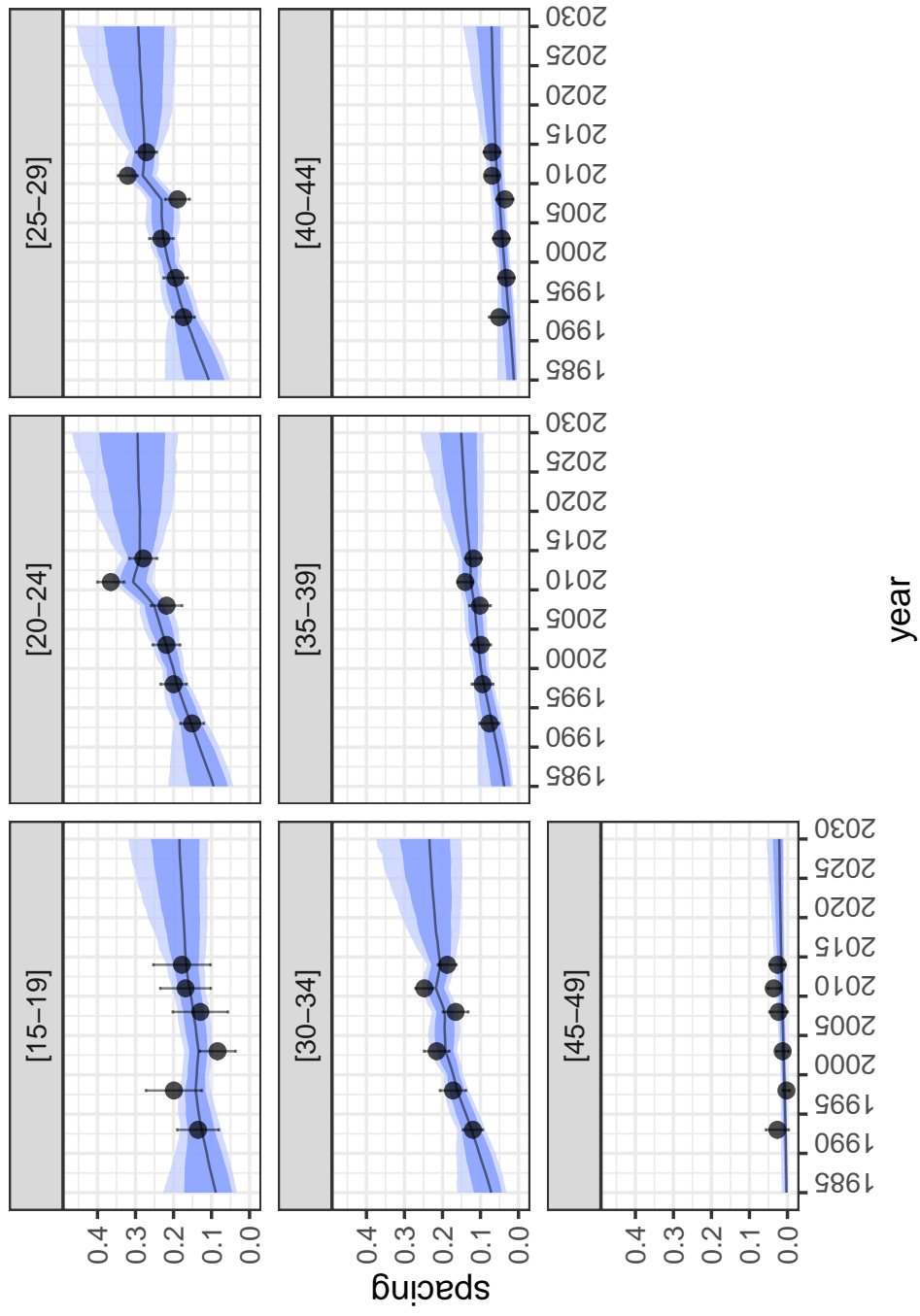
Gabon | Middle Africa



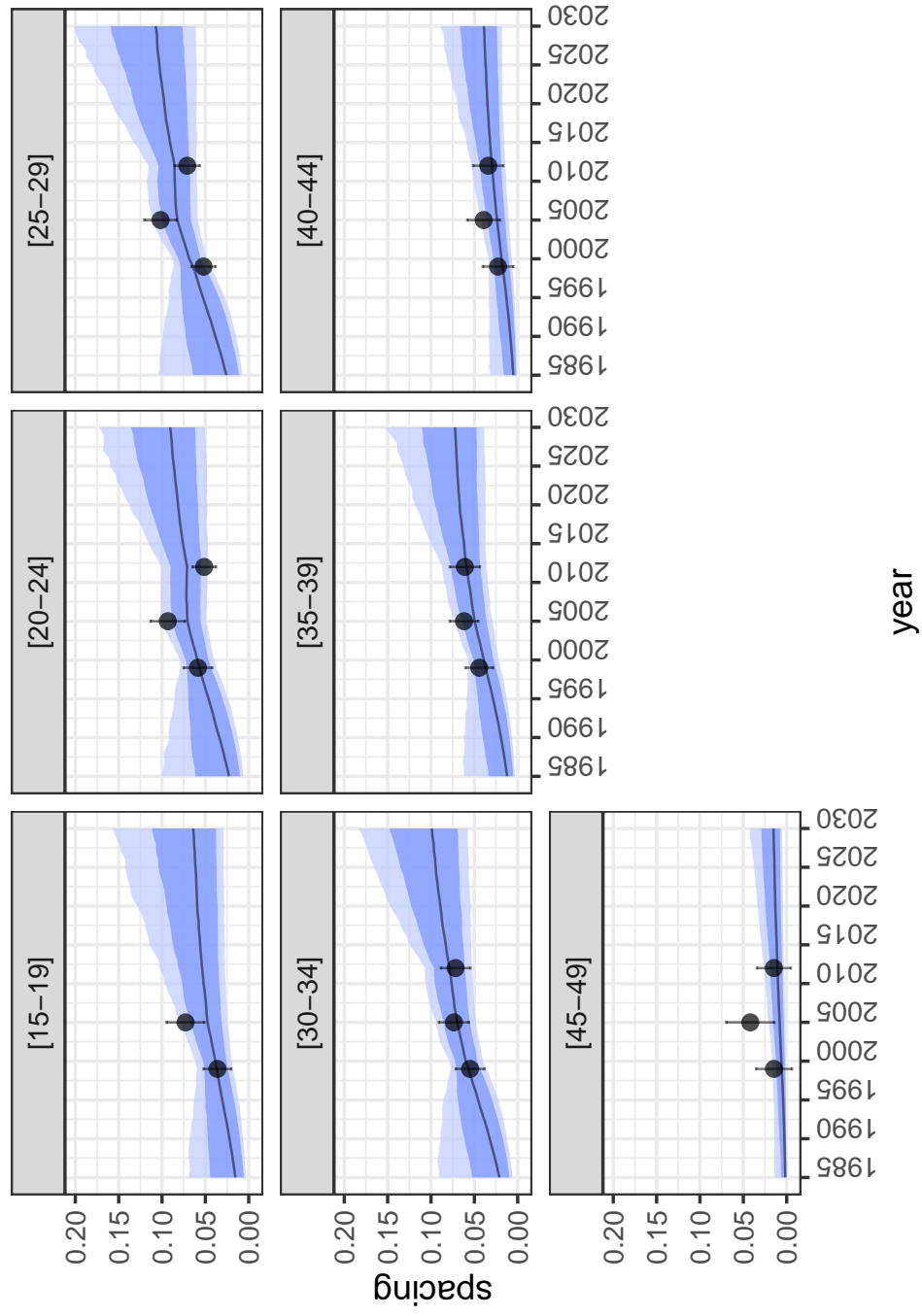
Gambia | Western Africa



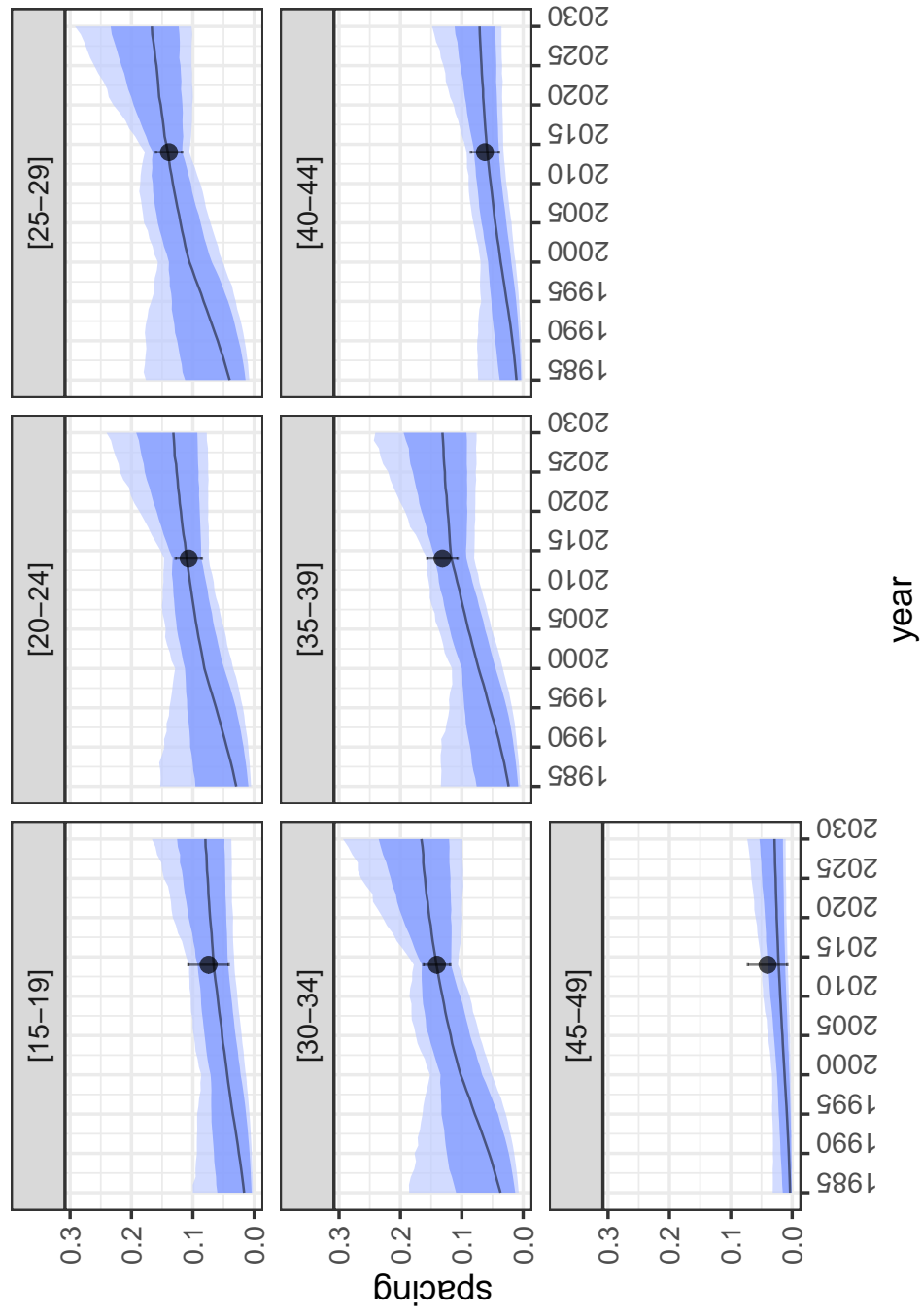
Ghana | Western Africa



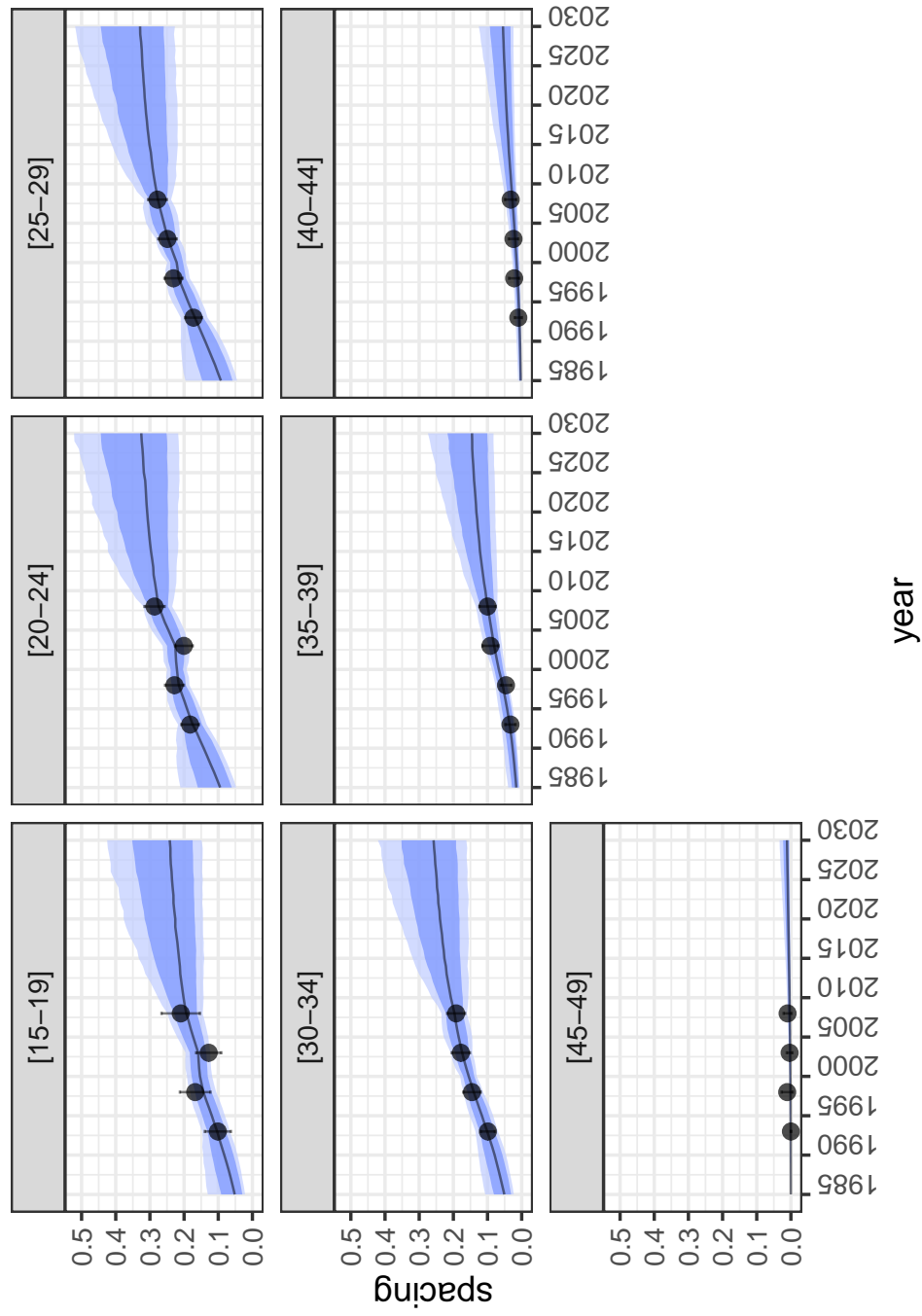
Guinea | Western Africa



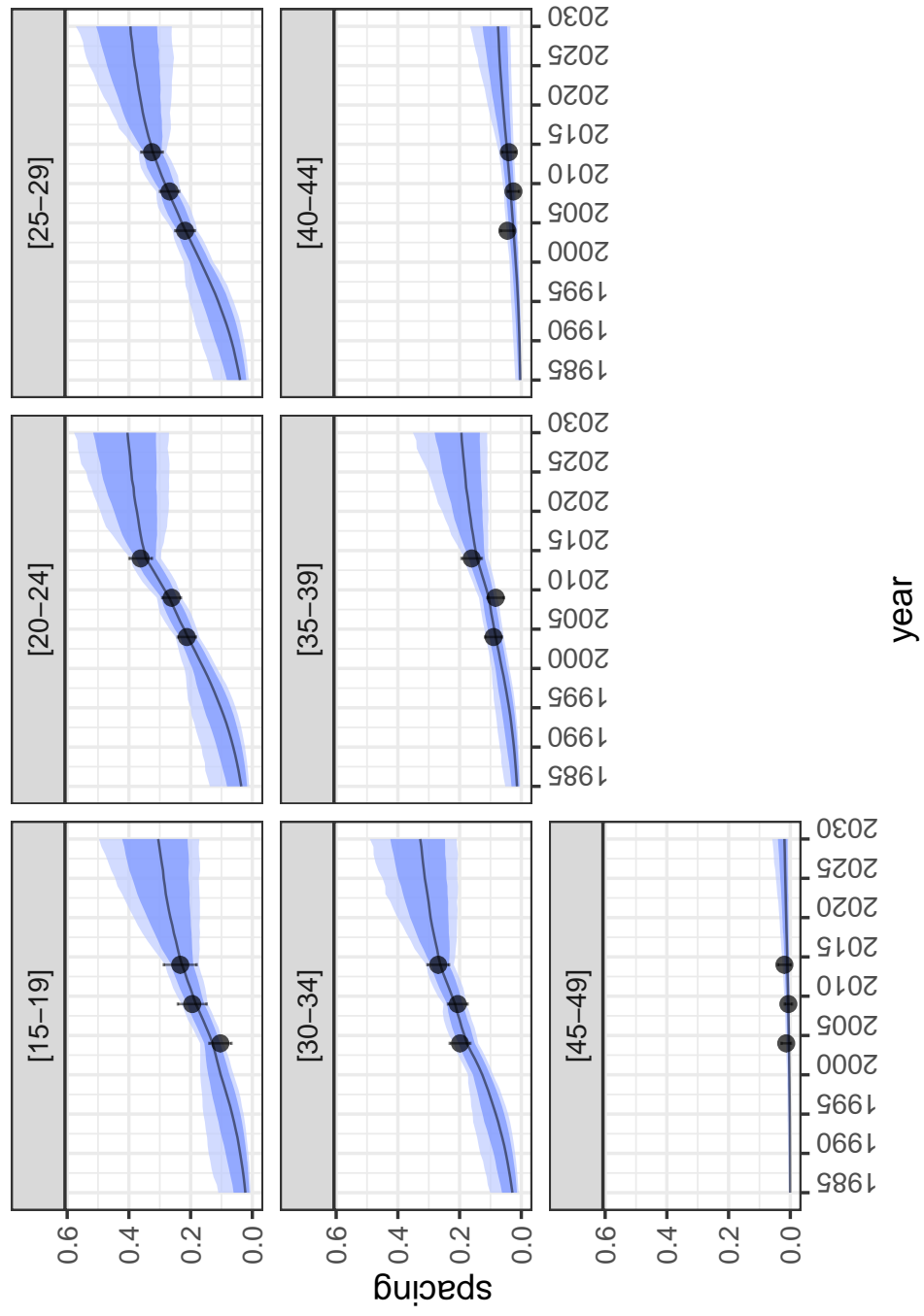
Guinea-Bissau | Western Africa



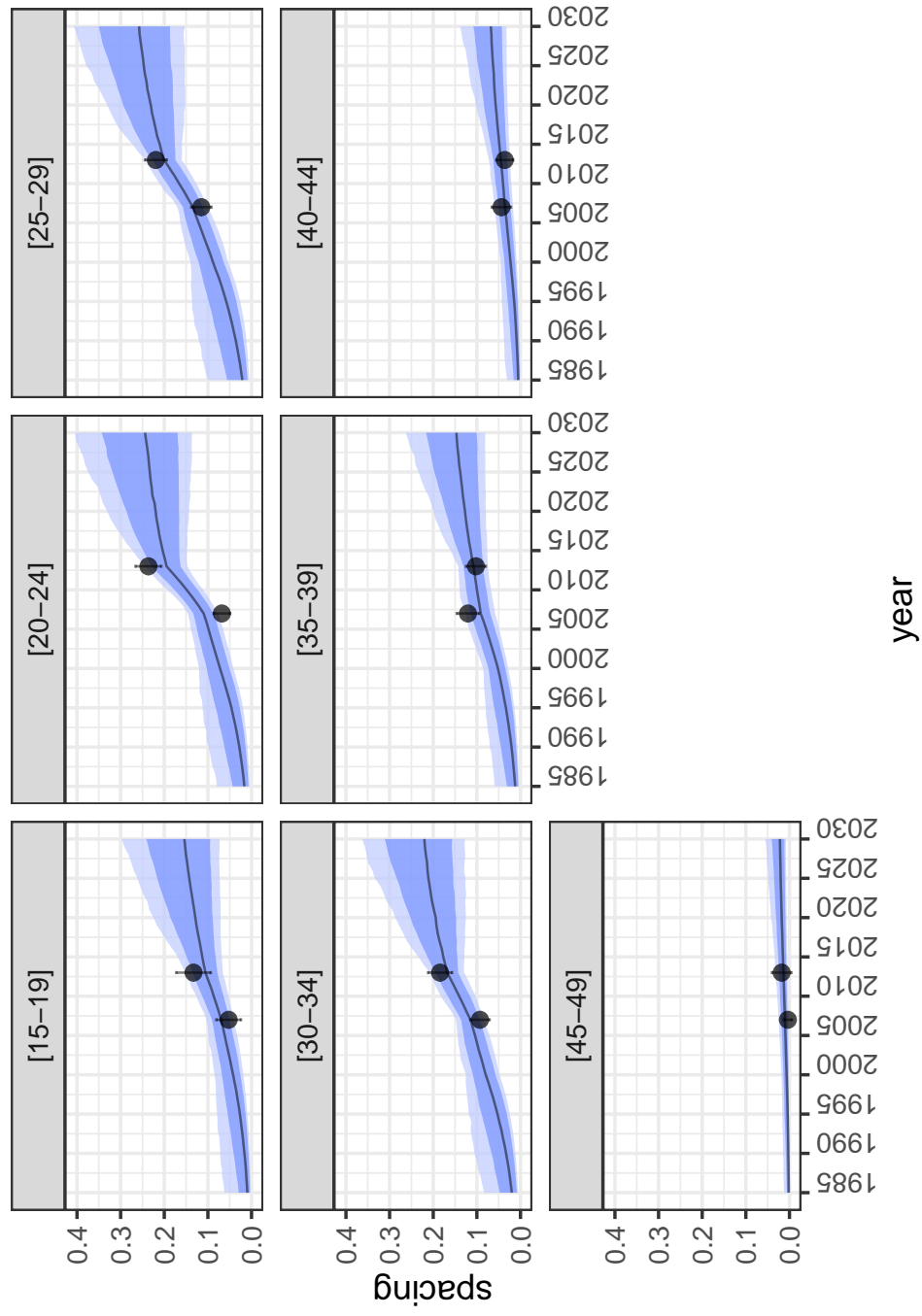
Kenya | Eastern Africa



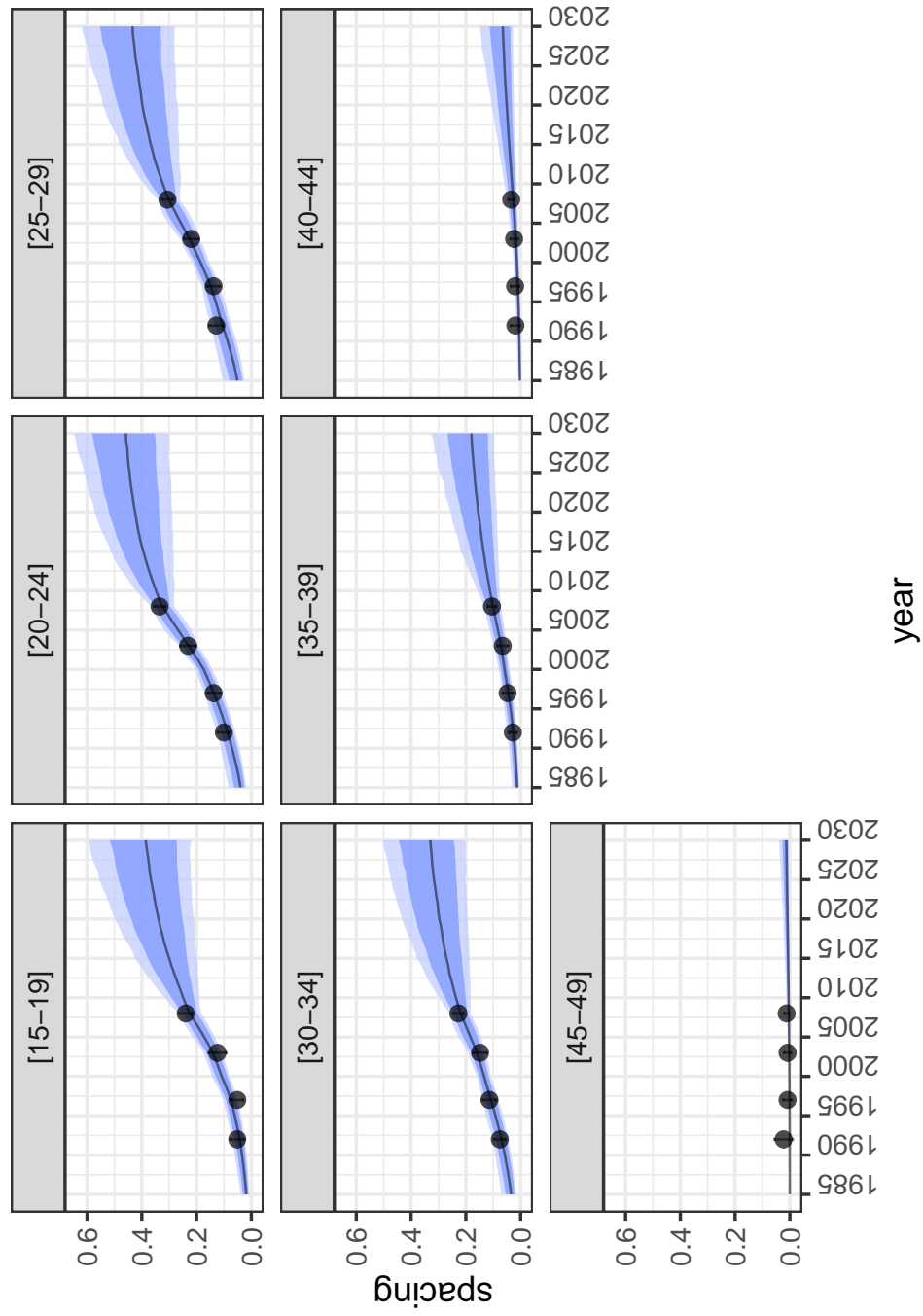
Lesotho | Southern Africa



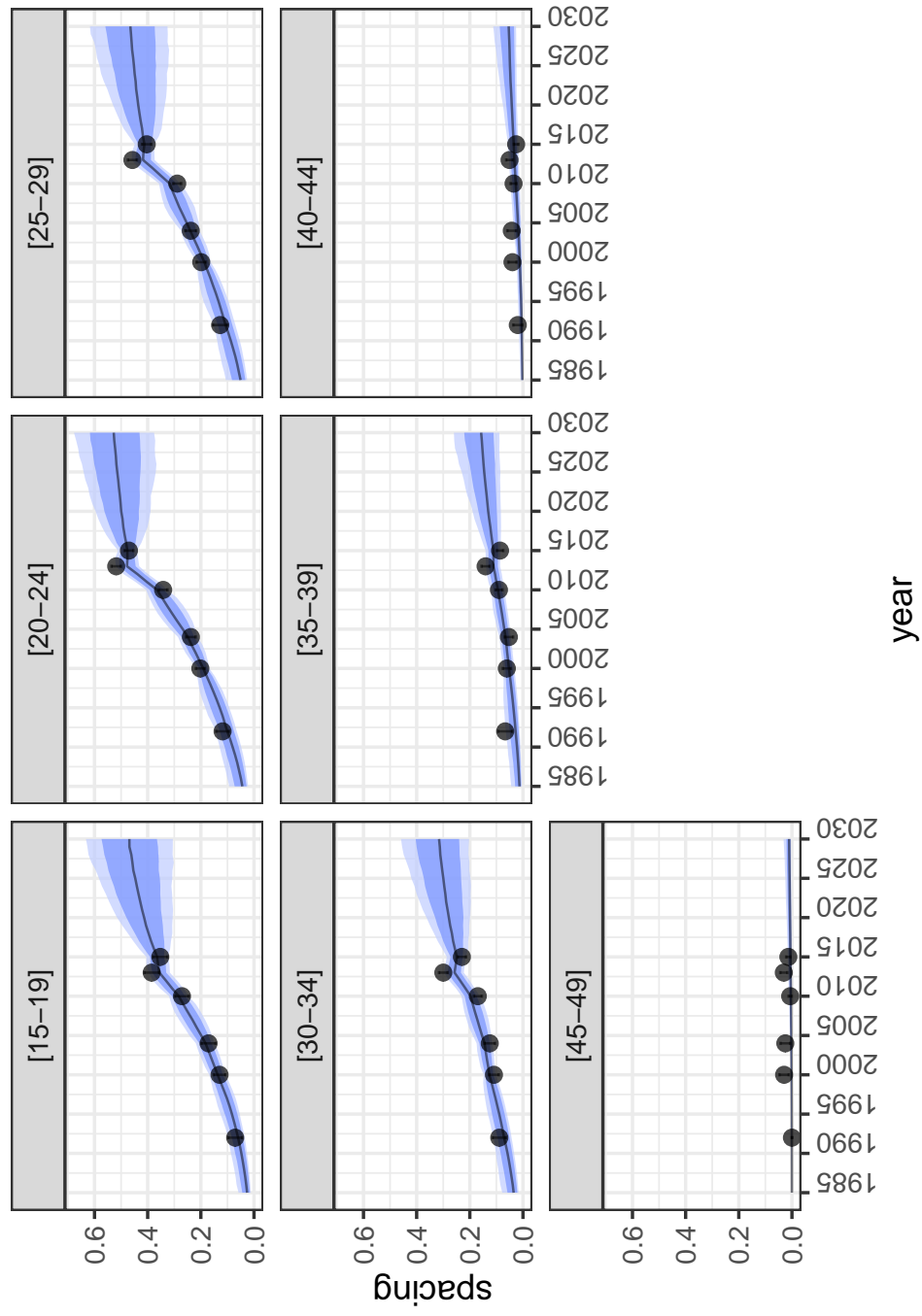
Liberia | Western Africa



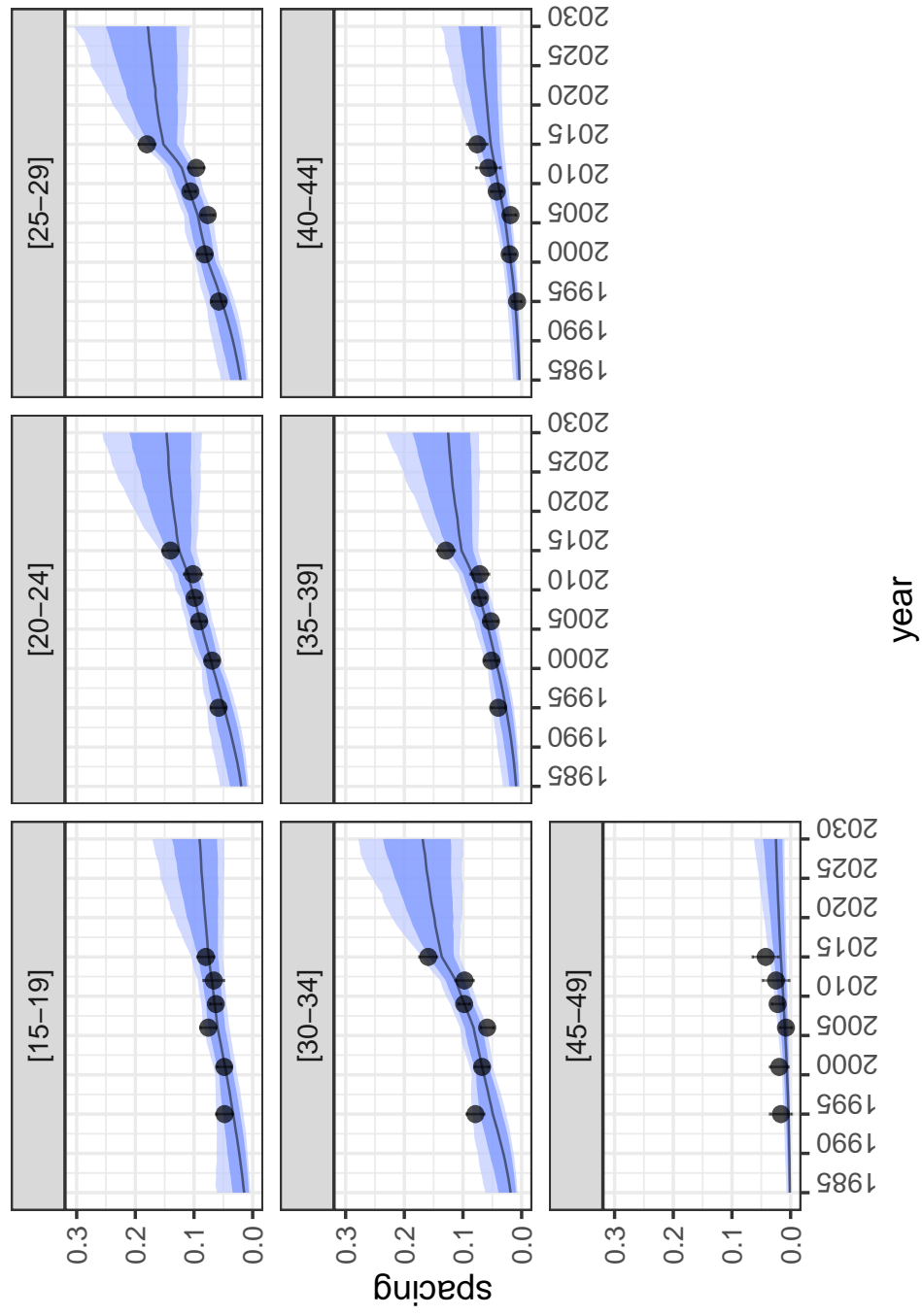
Madagascar | Eastern Africa



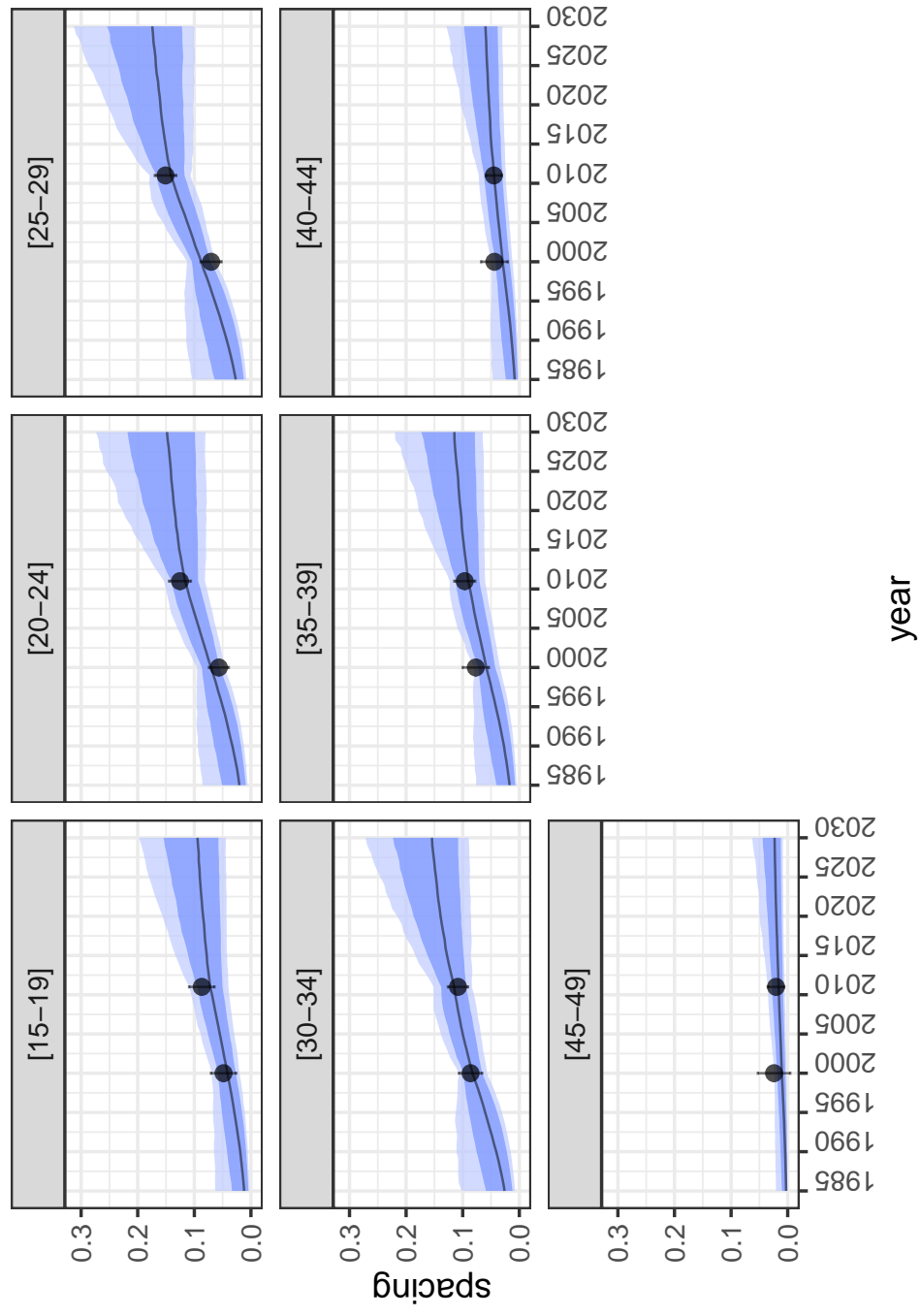
Malawi | Eastern Africa



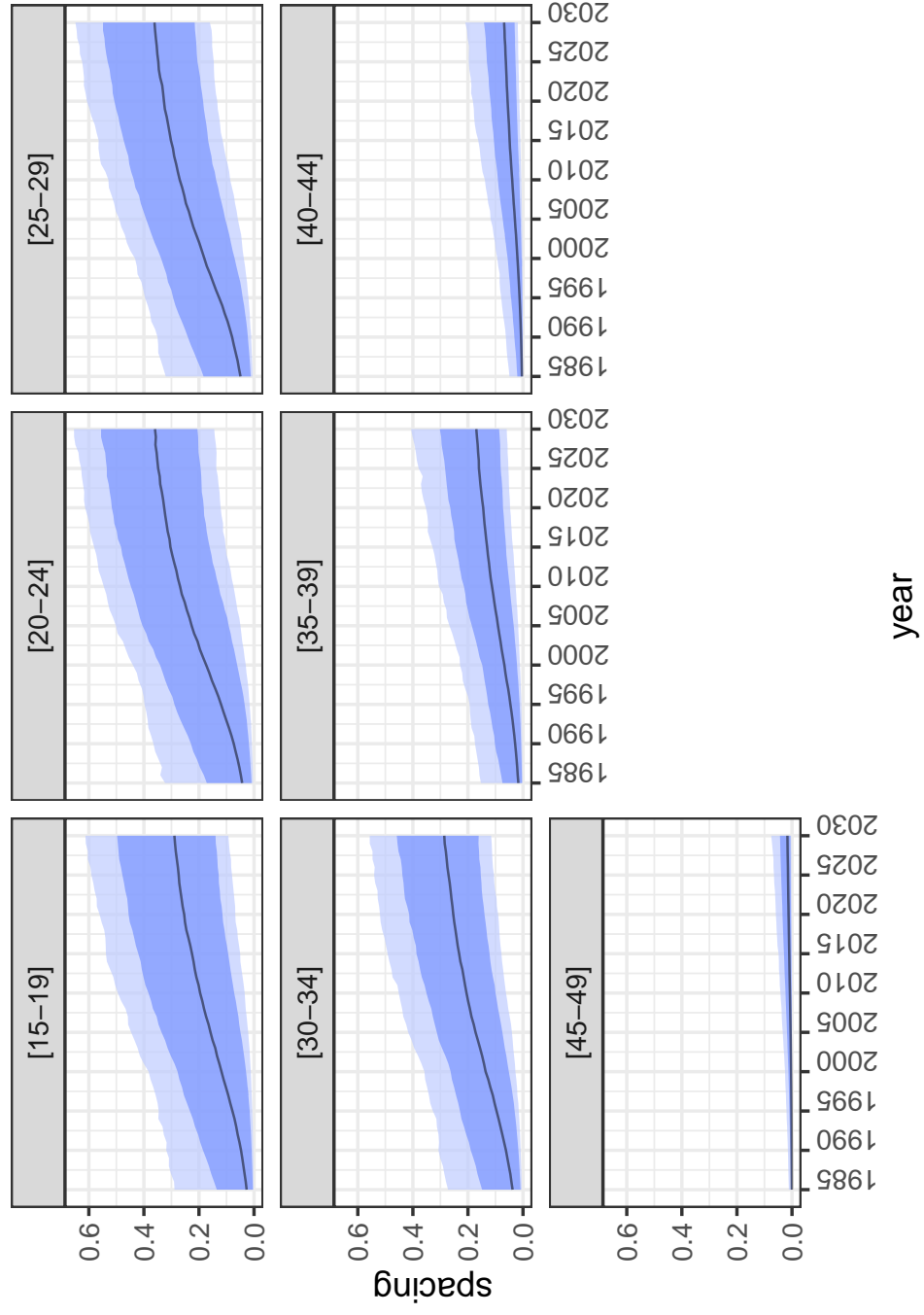
Mali | Western Africa



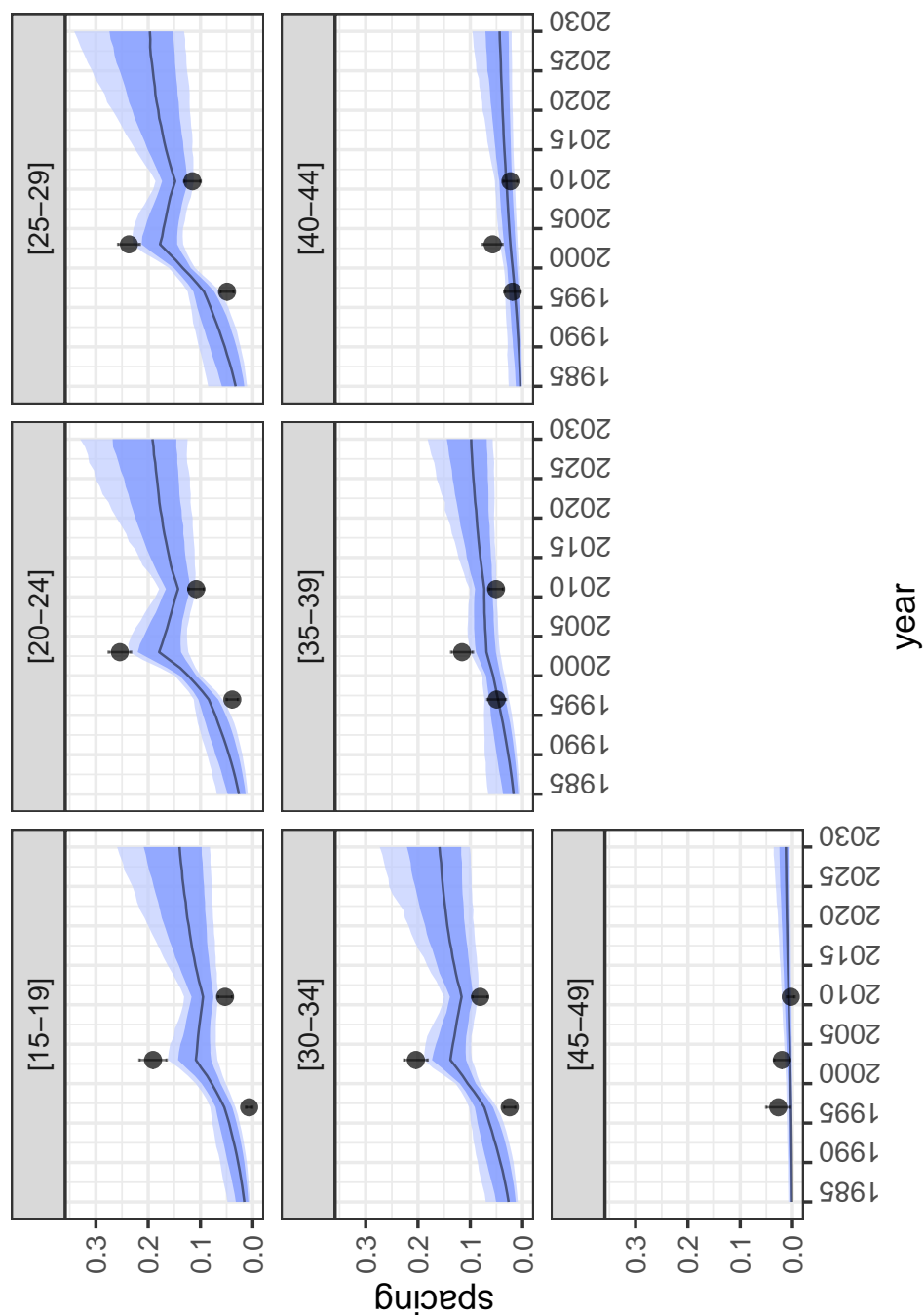
Mauritania | Western Africa



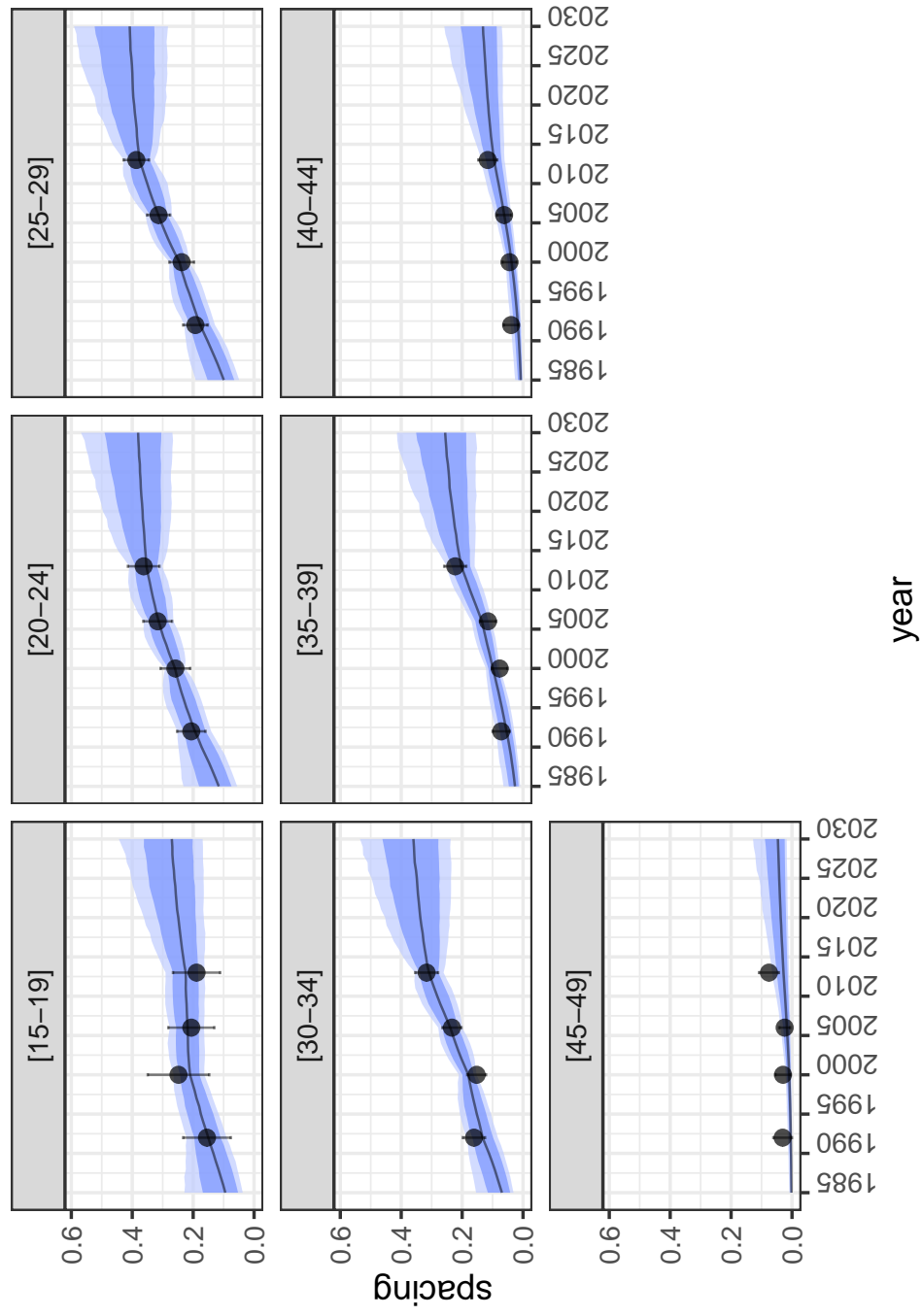
Mauritius | Eastern Africa



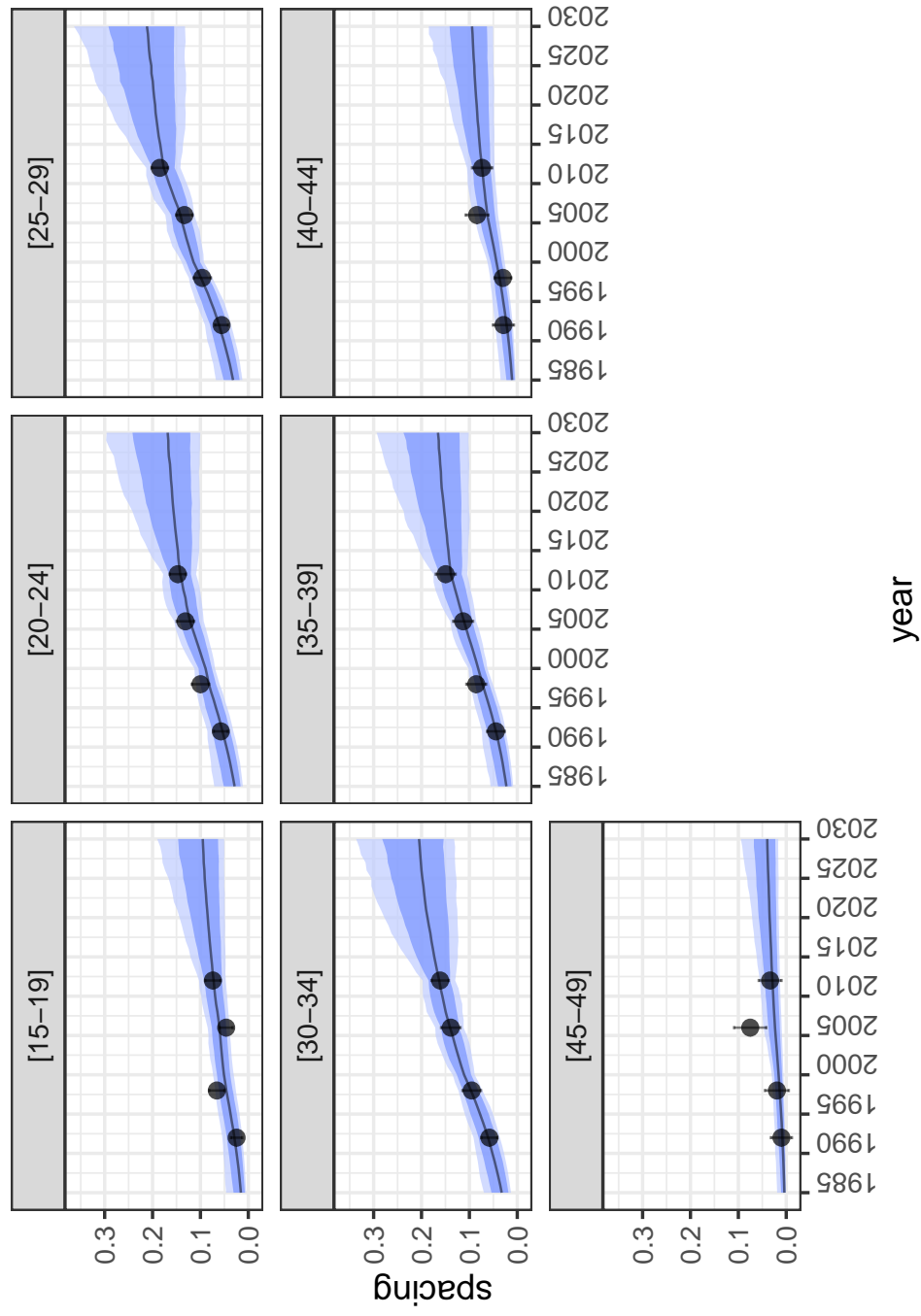
Mozambique | Eastern Africa



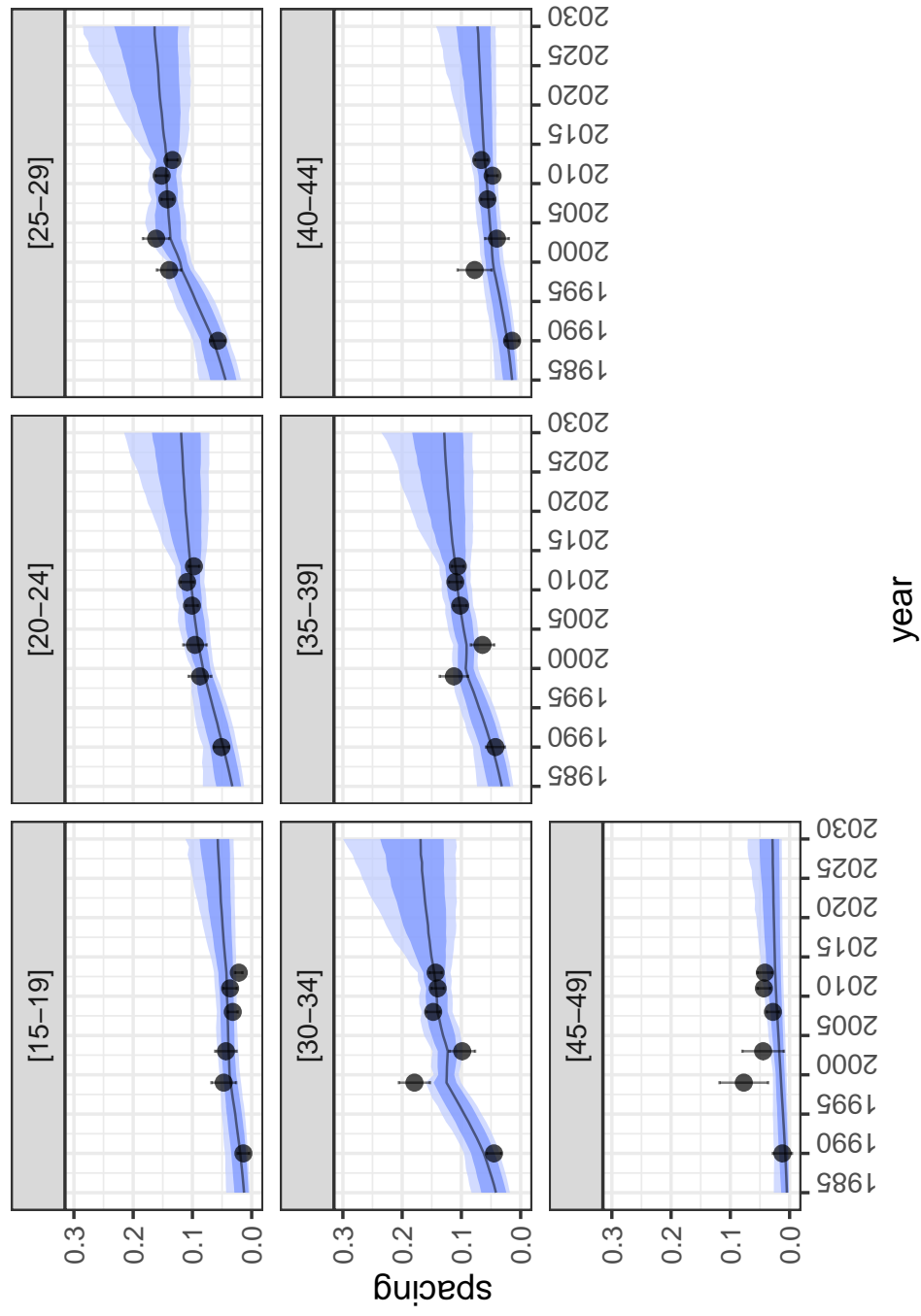
Namibia | Southern Africa



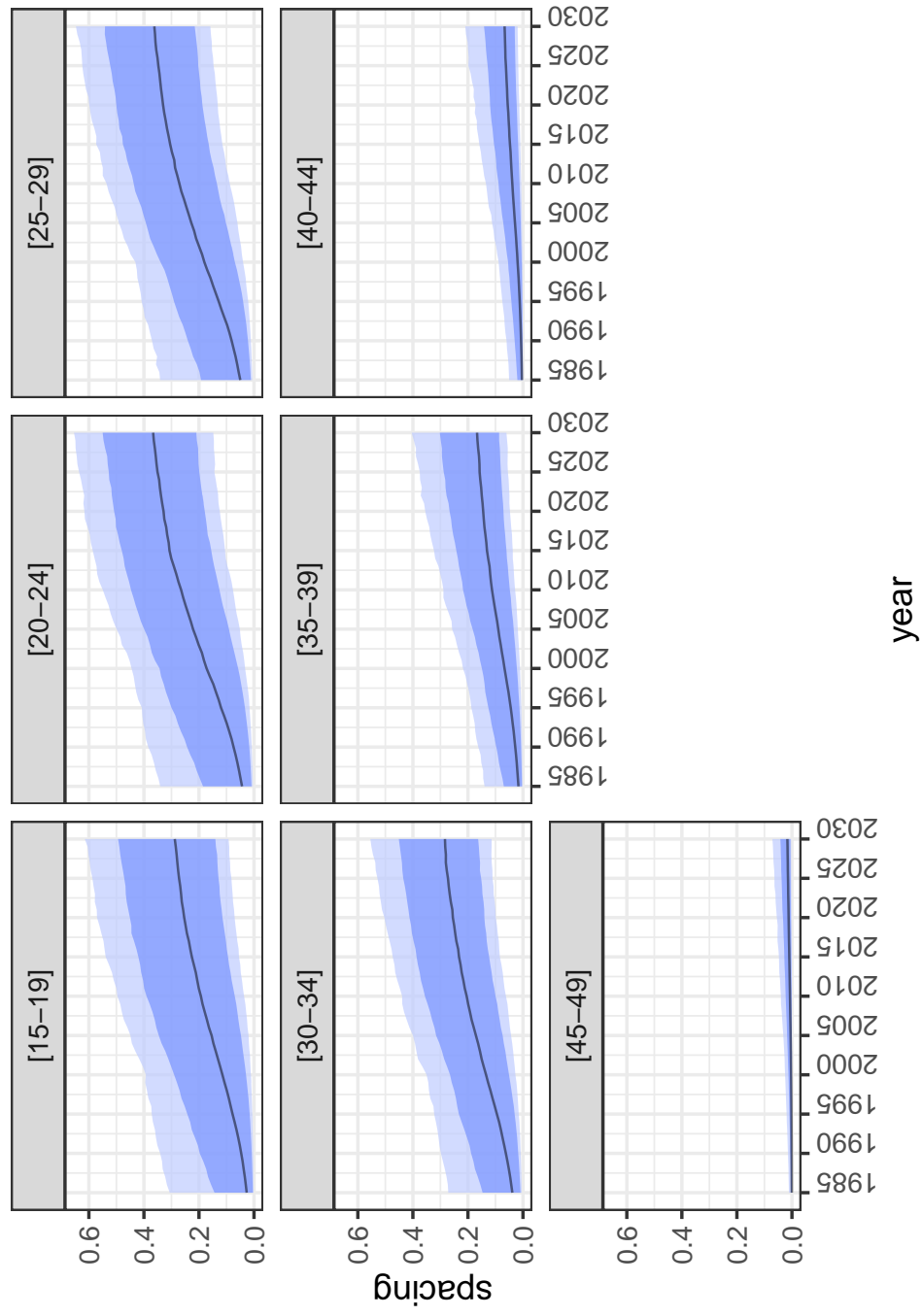
Niger | Western Africa



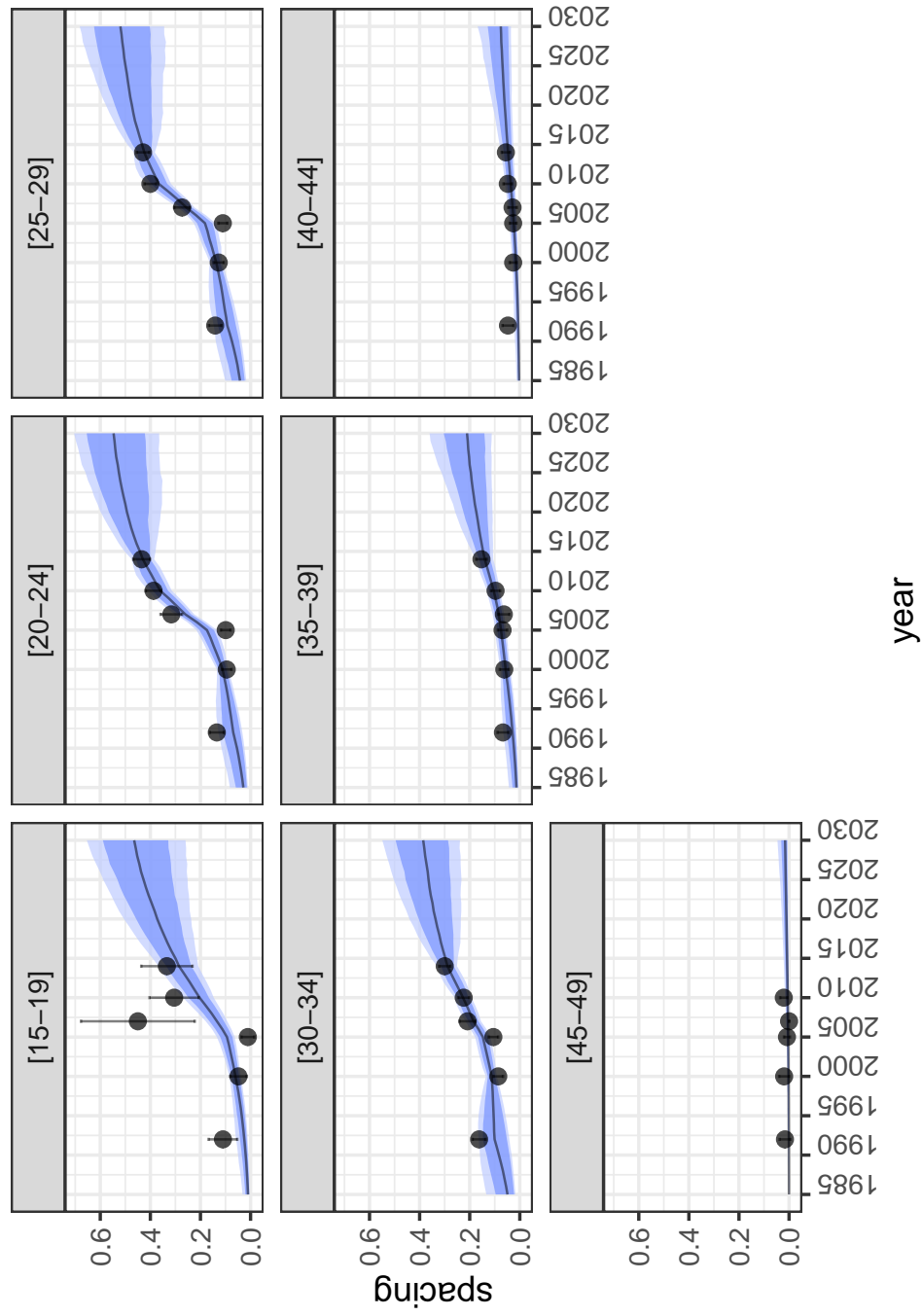
Nigeria | Western Africa



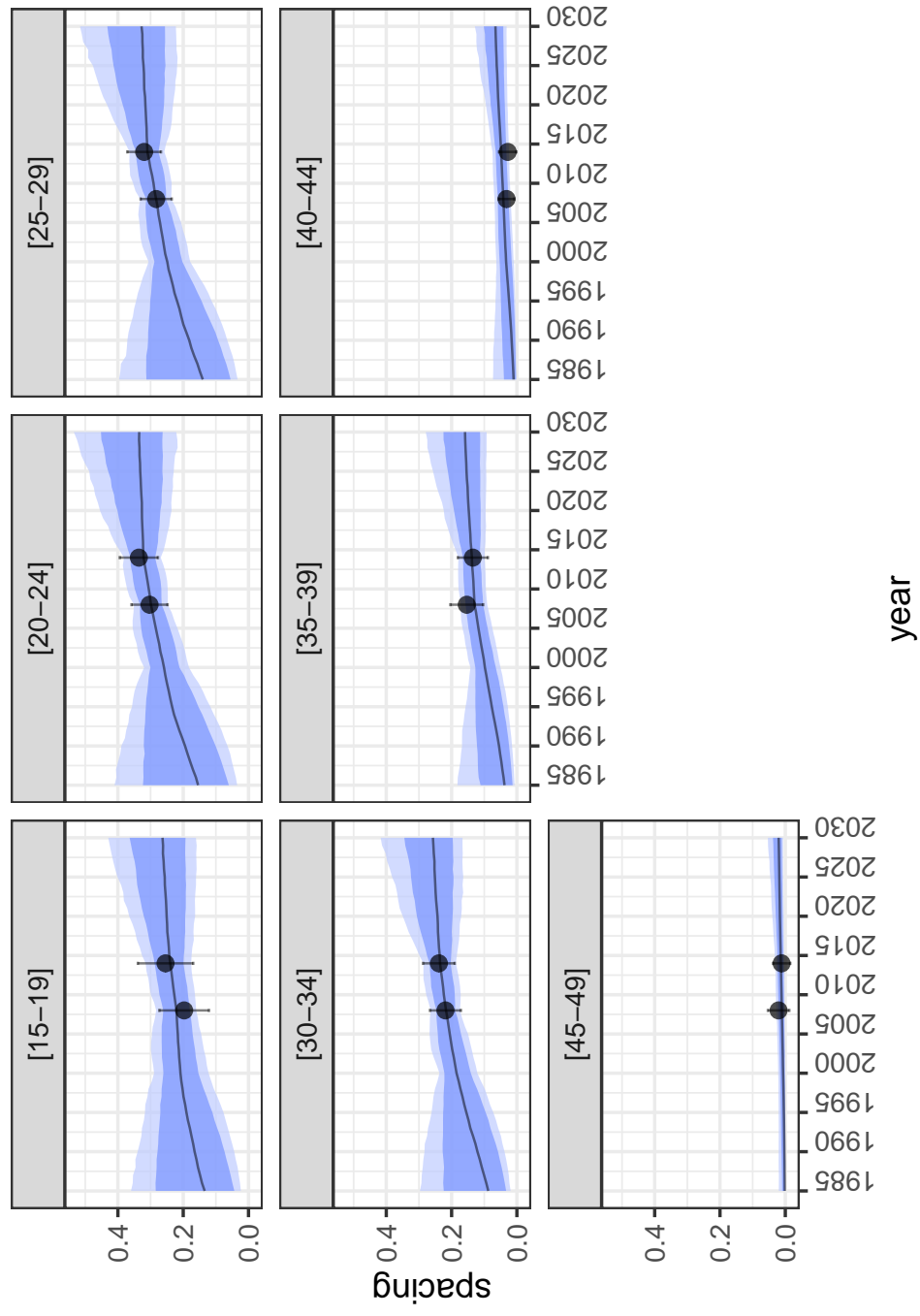
R.union | Eastern Africa



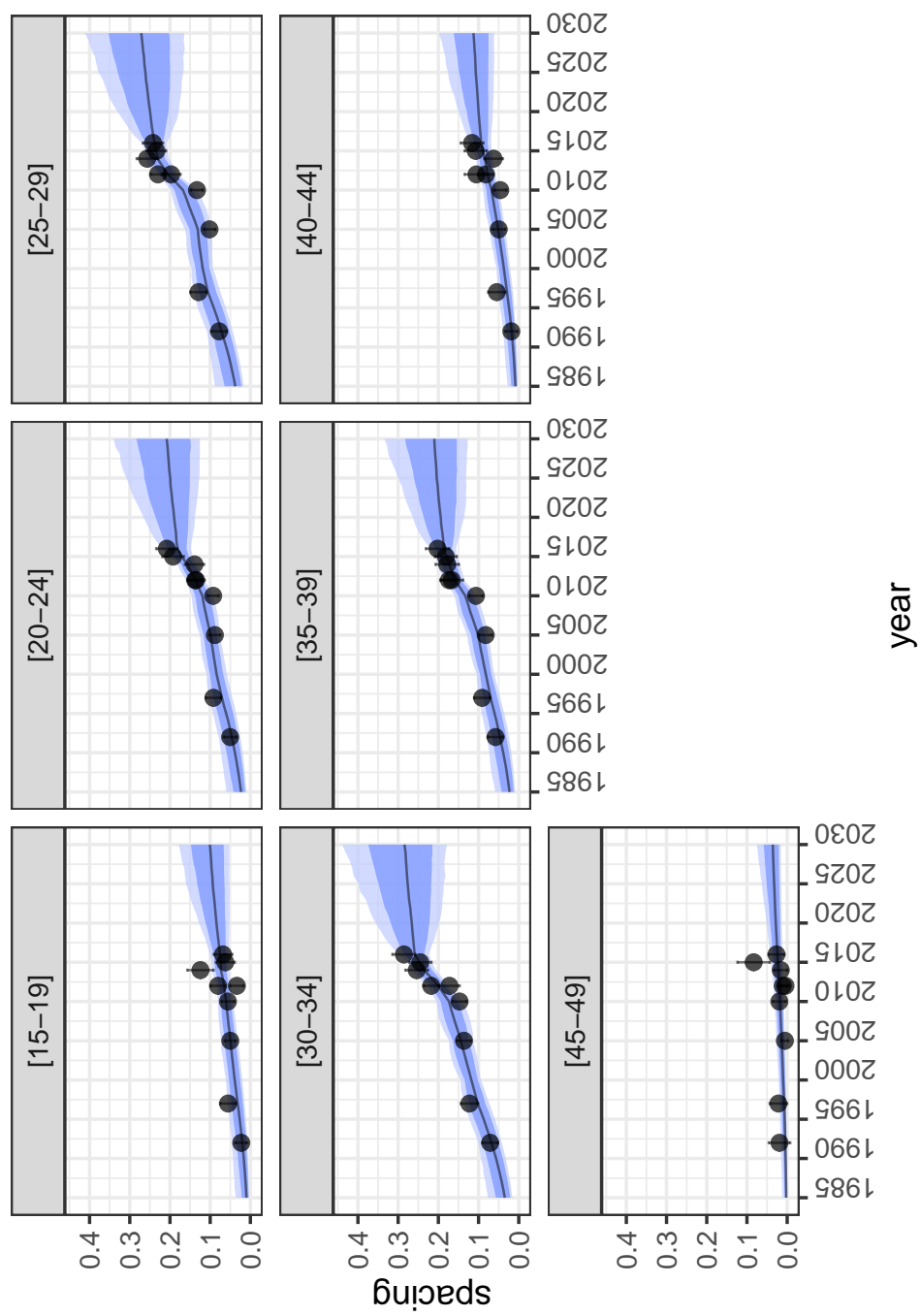
Rwanda | Eastern Africa



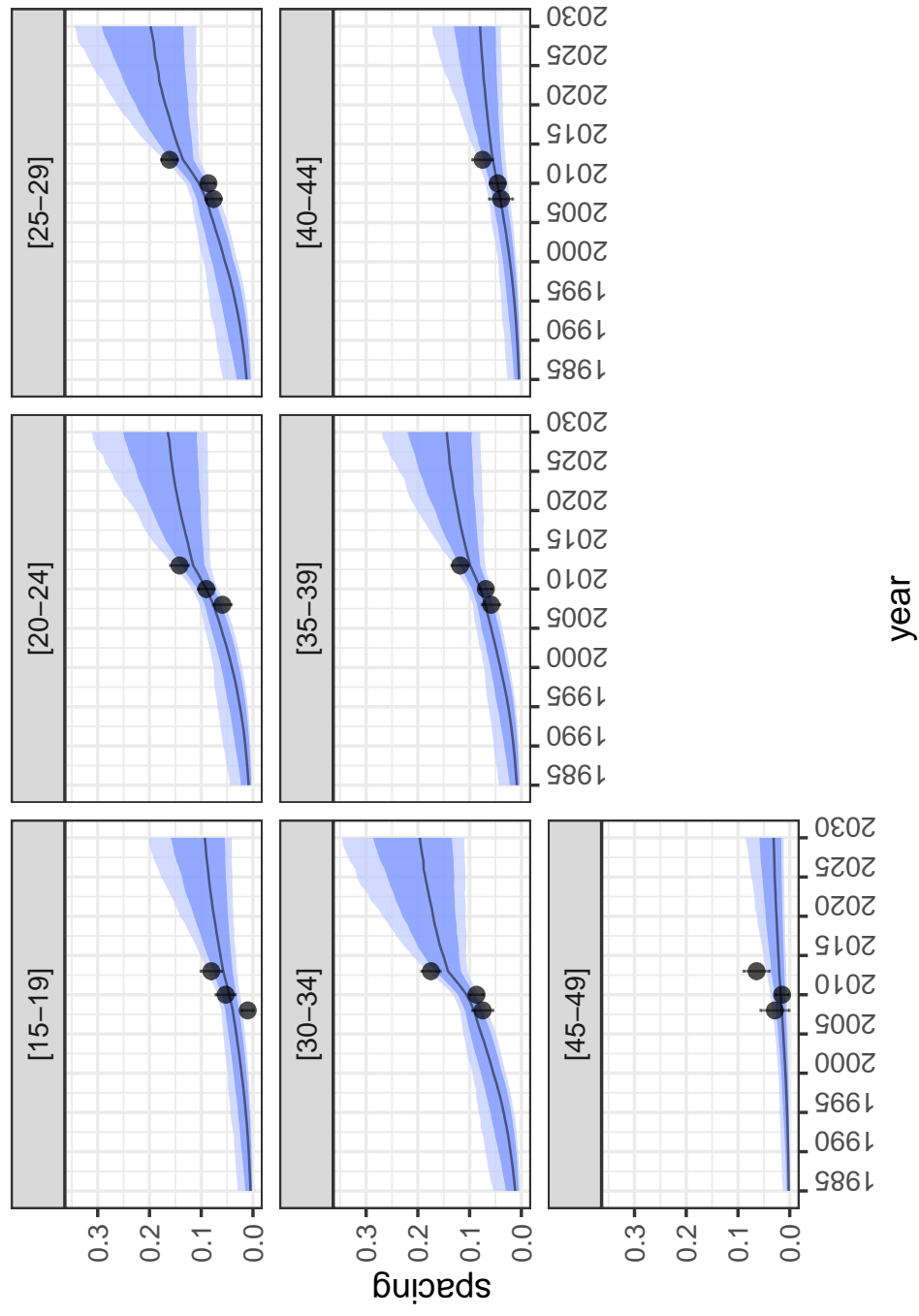
Sao Tome and Principe | Middle Africa



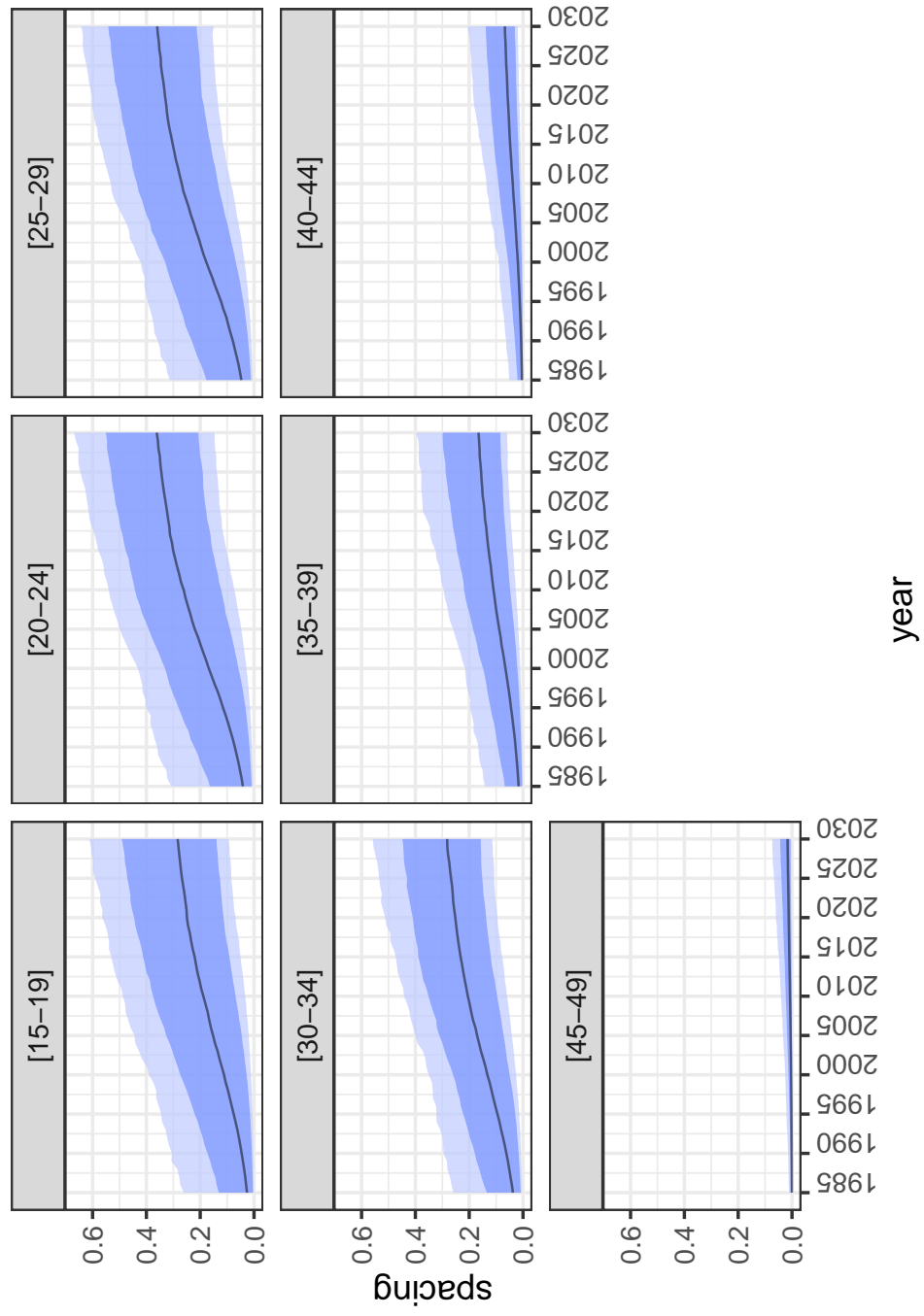
Senegal | Western Africa



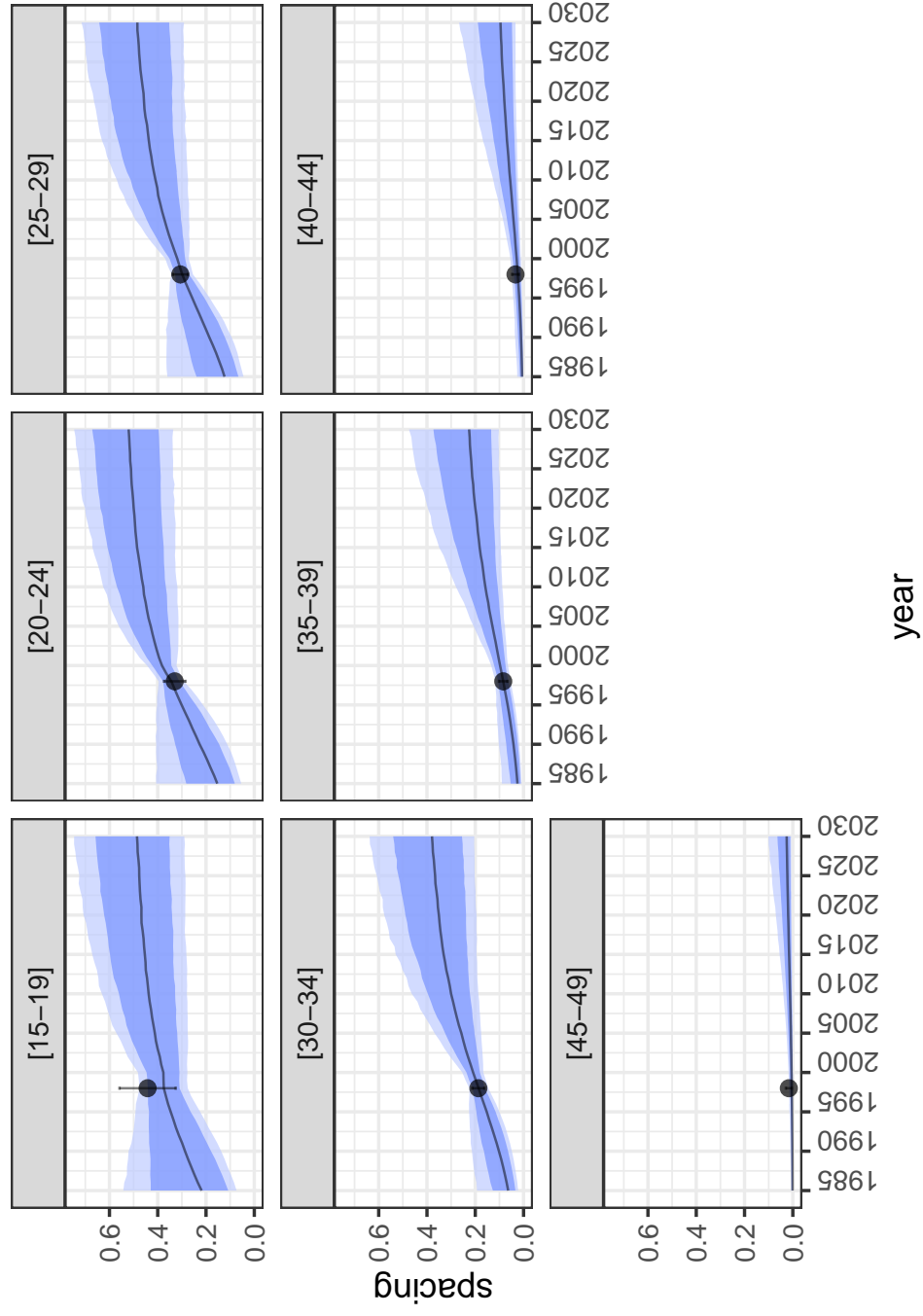
Sierra Leone | Western Africa



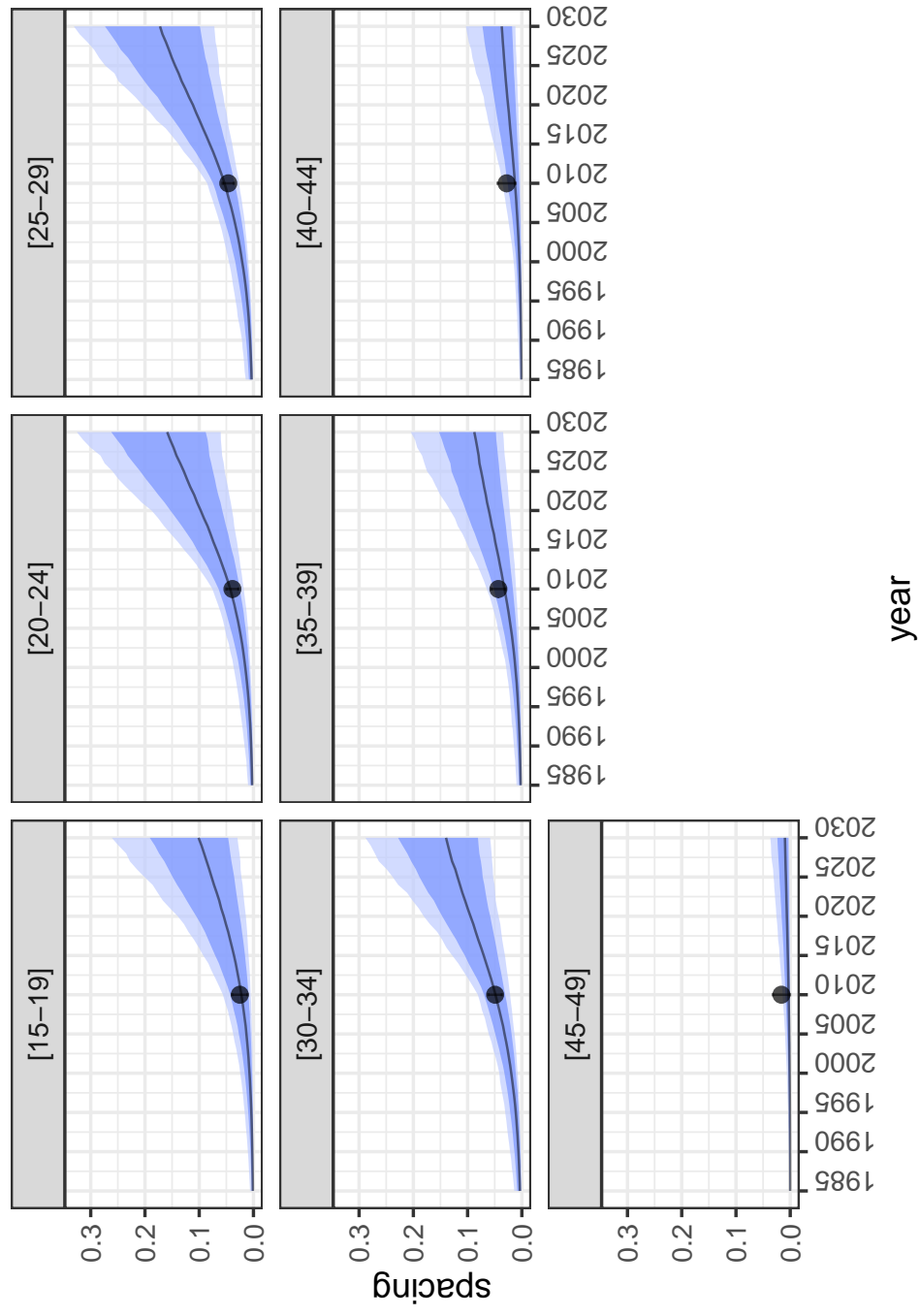
Somalia | Eastern Africa



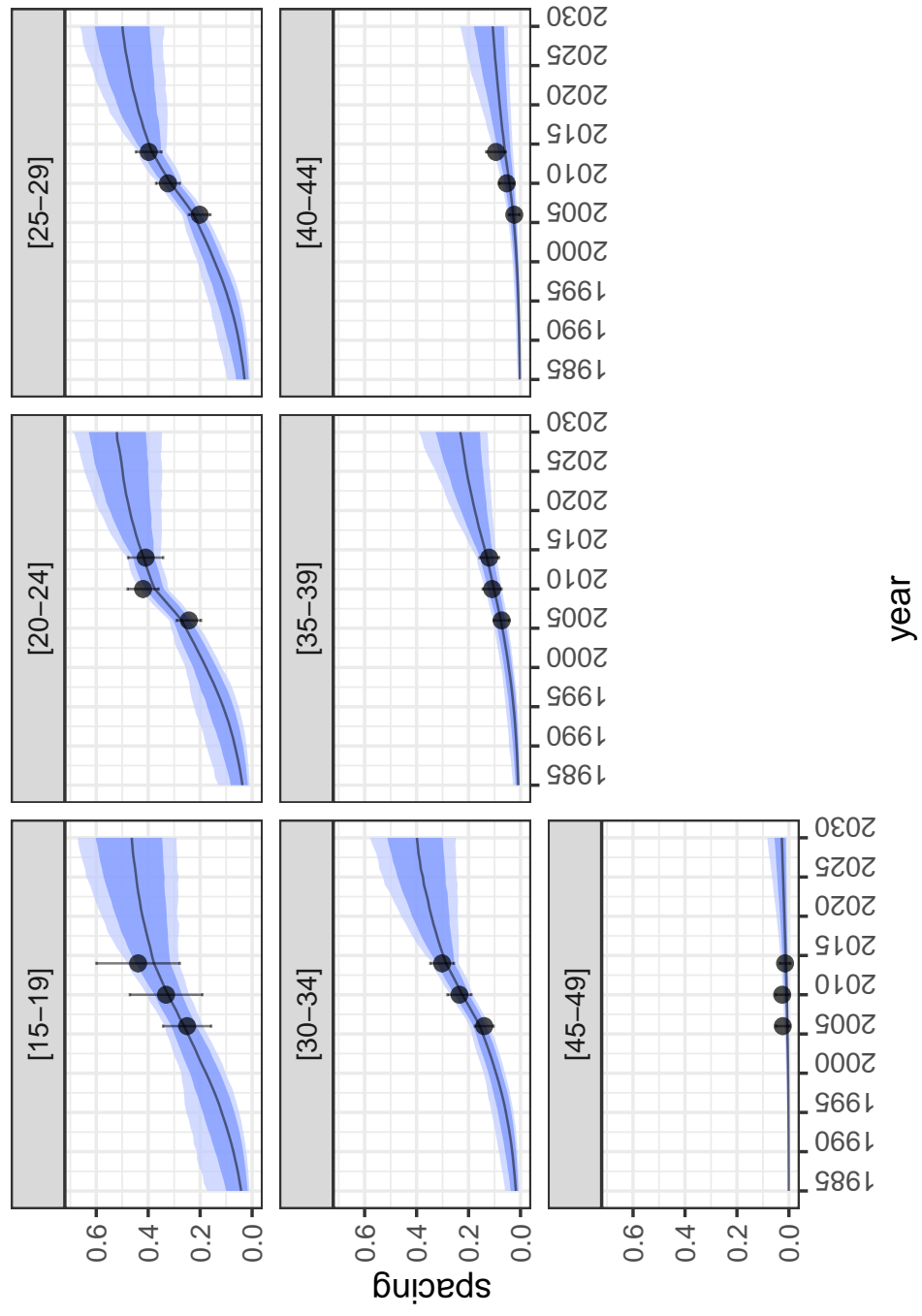
South Africa | Southern Africa



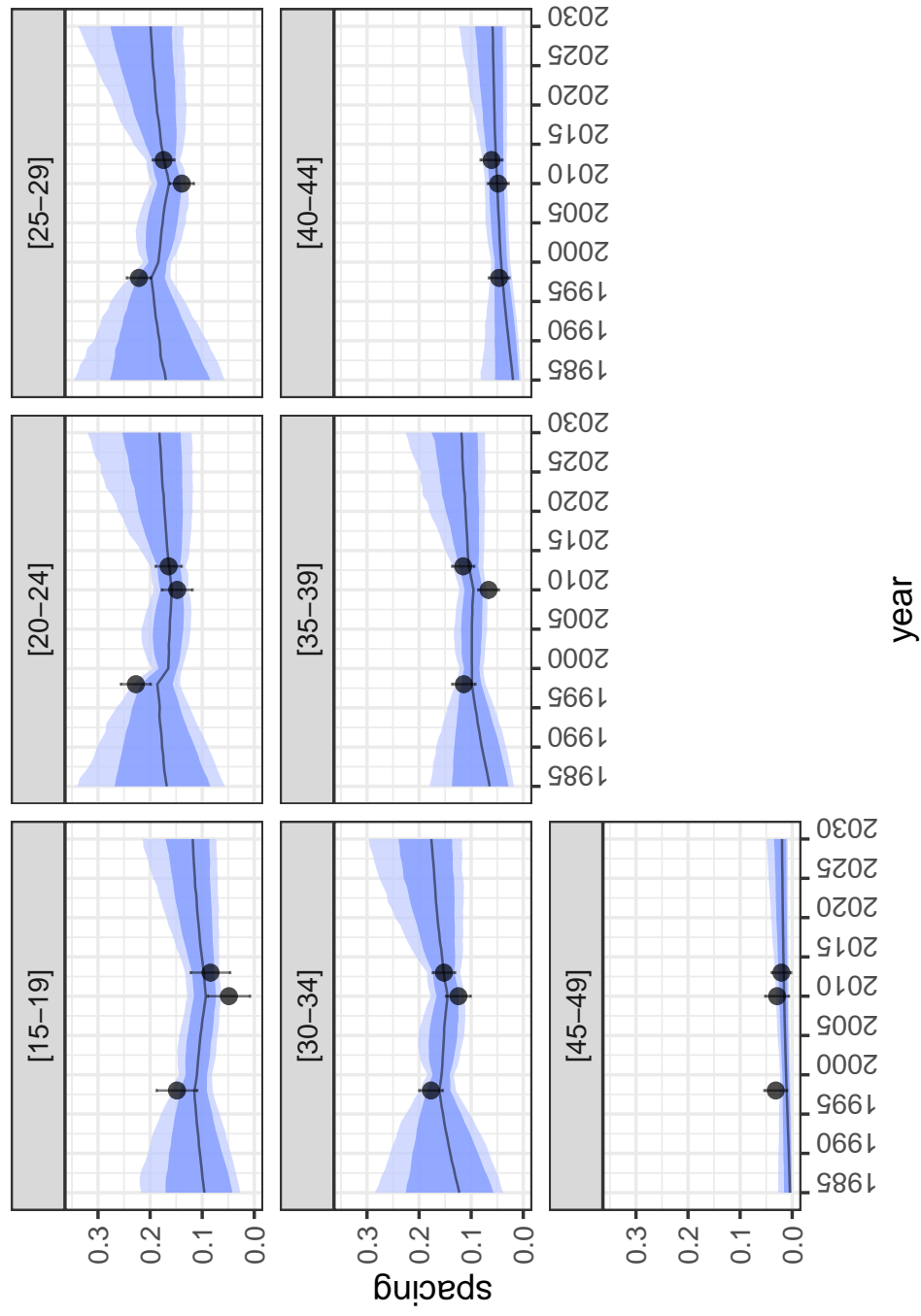
South Sudan | Eastern Africa



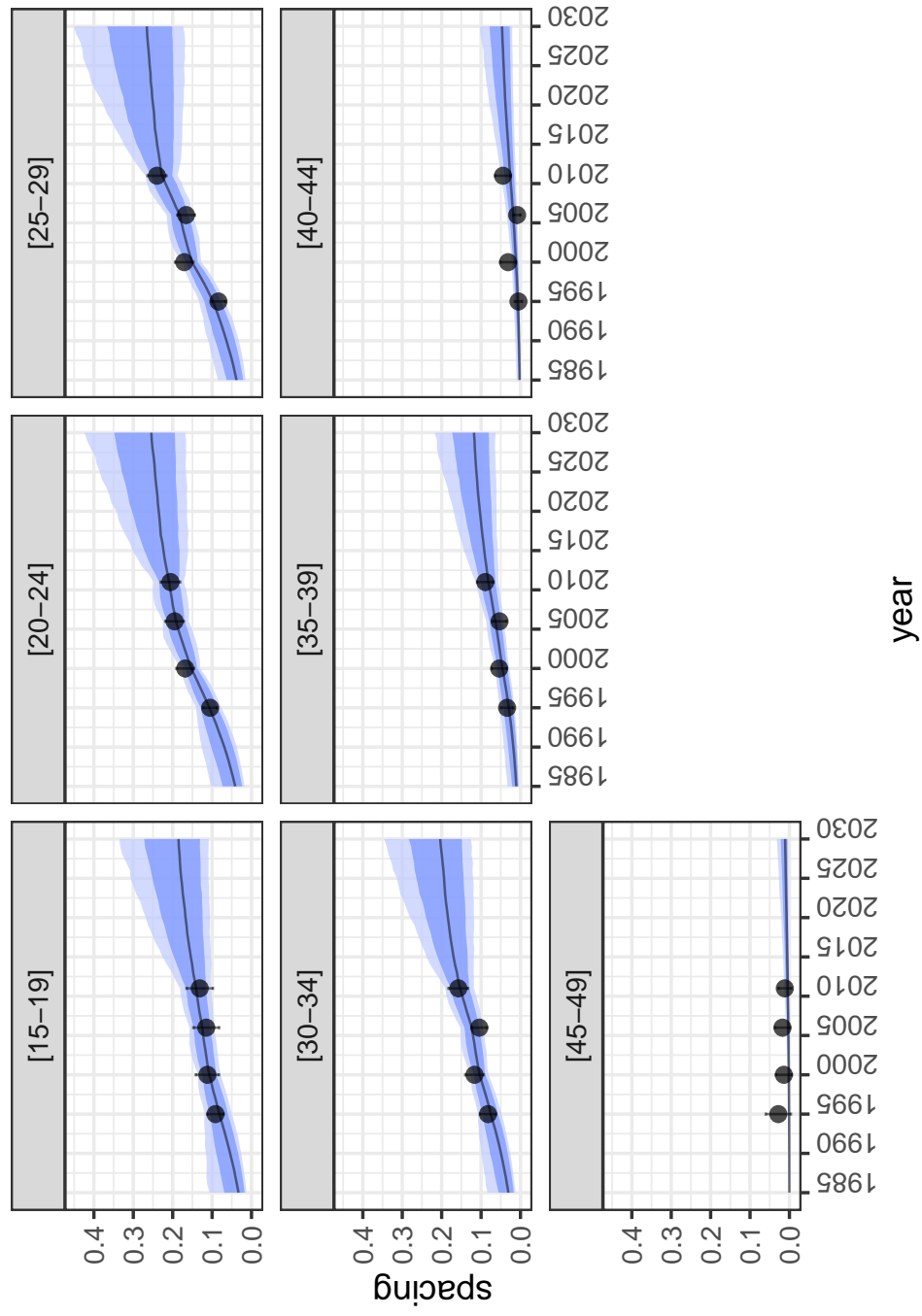
Swaziland | Southern Africa



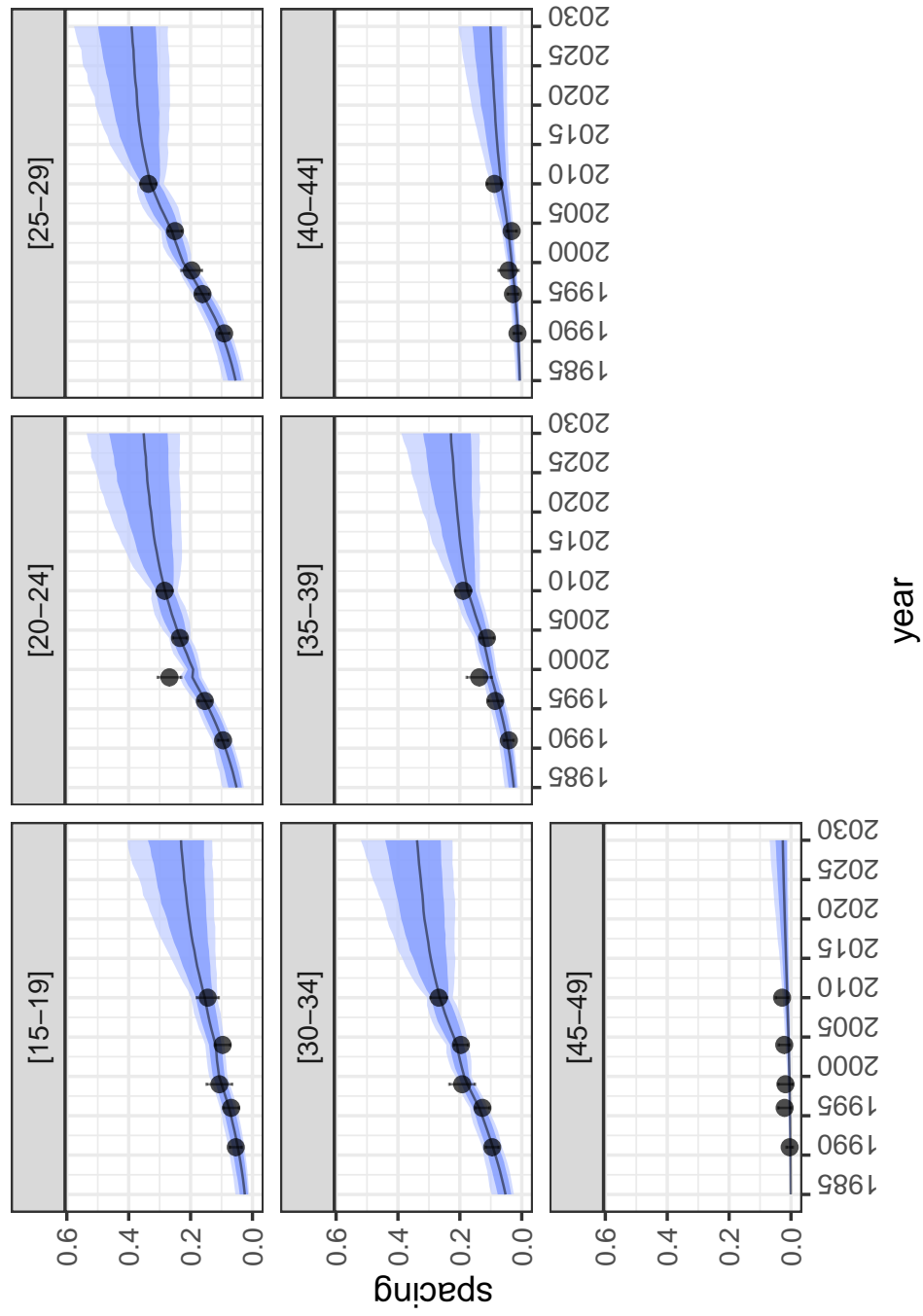
Togo | Western Africa



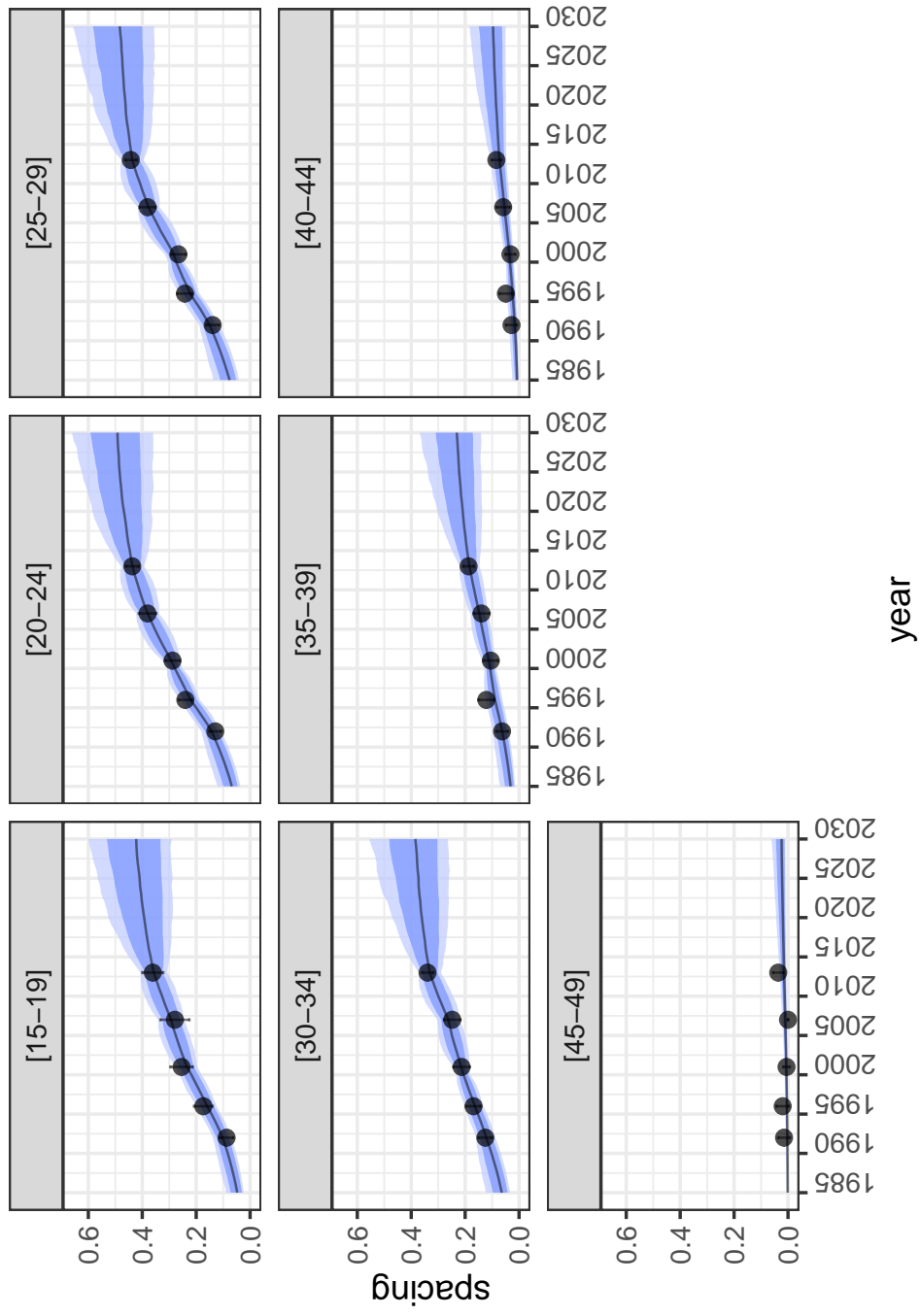
Uganda | Eastern Africa



United Republic of Tanzania | Eastern Africa



Zambia | Eastern Africa



BIBLIOGRAPHY

- Alkema, L., V. Kantorova, C. Menozzi, and A. Biddlecom (2013). National, regional, and global rates and trends in contraceptive prevalence and unmet need for family planning between 1990 and 2015: a systematic and comprehensive analysis. The Lancet (9878), 1642–1652.
- Bearak, J., A. Popinchalk, L. Alkema, and G. Sedgh (2018). Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a bayesian hierarchical model. The Lancet Global Health 6(4), e380–e389.
- Brown, W., N. Druce, J. Bunting, S. Radloff, D. Koroma, S. Gupta, B. Siems, M. Kerrigan, D. Kress, and G. L. Darmstadt (2014). Developing the 120 by 20 goal for the global fp2020 initiative. Studies in family Planning 45(1), 73–84.
- Cahill, N., E. Sonneveldt, J. Stover, M. Weinberger, J. Williamson, C. Wei, W. Brown, and L. Alkema (2018). Modern contraceptive use, unmet need, and demand satisfied among women of reproductive age who are married or in a union in the focus countries of the family planning 2020 initiative: a systematic analysis using the family planning estimation tool. The Lancet NA(10123), 870 – 882.
- DHS, M. (2006, September). Guide to dhs statistics.
- Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. Statistical Science 11(2), 89–121.
- Gelman, A. and J. Hill (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge university press.
- Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. Statistical Science 7(4), 457–472.
- JR New, L. A. Family planning estimation tool (fpet).
- Kantorová, V., J. R. New, A. Biddlecom, and L. Alkema (2017). Setting ambitious yet achievable targets using probabilistic projections: meeting demand for family planning. Studies in family planning 48(3), 223–233.

- Lindley, D. V. and A. F. M. Smith (1972). Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B 34, 1–41.
- Plummer, M. (2018). rjags: Bayesian Graphical Models using MCMC. R package version 4-8.
- Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In Proceedings of the 3rd international workshop on distributed statistical computing, Volume 124, pp. 10. Vienna, Austria.
- Plummer, M., N. Best, K. Cowles, and K. Vines (2006). Coda: convergence diagnosis and output analysis for mcmc. R news 6(1), 7–11.
- R Core Team (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Singh, S., G. Sedgh, and R. Hussain (2010). Unintended pregnancy: worldwide levels, trends, and outcomes. Studies in family planning 41(4), 241–250.
- Su, Y.-S. and M. Yajima (2015). R2jags: Using R to Run 'JAGS'. R package version 0.5-7.
- United Nations, D. o. E. and P. D. Social Affairs (2017). World contraceptive use 2017.